Coupling time-indexed and big-M formulations for real-time train scheduling during metro service disruptions

Yeran Huang[1], Carlo Mannino[*2], Lixing Yang[*1], and Tao Tang[1]

[1]*State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, 100044, China*
[2]*Department of Mathematics and Cybernetics, SINTEF Digital, Oslo, and University of Oslo*

Abstract

Track disruptions in metro systems may lead to severe train delays with many passengers stranded at platforms, unable to board on overloaded trains. Dispatchers may put in place different recovery actions, such as alternating train directions and allowing short turns. The objective is to alleviate the inconvenience for passengers and to regain the nominal train regularity. To characterize this process, this paper develops nonlinear mixed integer programming (NMIP) models with two different recovery strategies to reschedule trains during the disruption. For solving models in real time, the hybrid formulation, which couples big-M and time-indexed formulations, is proposed to linearize the proposed model as the mixed integer linear programming (MILP) model. Then, a two-stage approach is designed for handling the real-time detected information (like dynamic arriving passengers and end time of the disruption), including offline task (to select the best recovery strategy) and online task (to implement the best strategy and update timetable). Finally, the numerical experiments from Beijing metro Line 2 are implemented to verify the performance and effectiveness of the proposed hybrid formulation and two-stage approach.

Keywords: Metro disruption; Dynamic arriving passengers; Recovery strategies; Two-task approach

1   Introduction

With the impressive expansion of the cities, urban transportation systems are under pressure everywhere. For instance, in 2018, the passenger flow in Beijing metro network almost exceeded 10 million persons per day [2], which accounts for 50% of the total flow in the Beijing public transit system [27]. A crucial role in subway systems is played by ring lines such as Beijing Lines 2 and 10, Shanghai Line 4, Chengdu Line 7, etc. These lines commonly orbit the central business district and connect with radial lines, bearing a large share of the overall transport pressure. For instance, the Beijing metro system consists of 20 lines, but the ring Lines 2 and 10 together account for approximately 20% of the total passenger flow [2]. Indeed, transport demand and consequent pressure over the system is increased. In peak hours, in some lines of Beijing metro, the headway between two consecutive trains has been shortened to 2 minutes; nevertheless, the average load rate of trains exceeds 140% [36]. This rate means that trains run overcrowded with passengers pressed against each other.

---

*Corresponding authors. E-mail addresses: carlo.mannino@sintef.no (C. Mannino), lxyang@bjtu.edu.cn (L. Yang)

In practice, service disruptions frequently occur, with partial blockage of the urban rail transit. Disruptions may be caused by different reasons, such as rolling stock failure, track failure, intrusions, medical emergencies, weather/nature disasters, etc. [26, 34]. The blockage induces service train delay, with passengers stranded at stations unable to board on trains. For instance, in two months of 2017 in the Beijing Subway, a total of 8 incidents occurred - including 2 passenger intrusions, 1 rolling stock failure and 5 track failures - with a service delay varying from 15 to 40 minutes.

In the recovery management process under a disruption scenario, we identify three successive phases with different impacts on the system's capacity, as illustrated in Figure 1 (see [20, 14]). In Phase I after the incident occurs, the traffic capacity declines dramatically and very quickly from the normal service level. Dispatchers identify the causes of the disruption and estimate its duration. Additionally, the recovery strategy is established, including train rescheduling, passenger flow guidance, and bus bridging service. Phase II begins when the established recovery actions are finally put in place. Capacity starts rising again at a pace that depends on the effectiveness of the implemented actions. Phase III starts when the disruption is over. After some time, the normal traffic capacity is regained.

From this schematic phasing, the crucial role played by the recovery actions in regaining the normal regime in a short time is apparent. In practical operations, a number of train rescheduling strategies are usually adopted, such as holding trains, short-turn operations, adding gap trains, station skipping, etc. (see [34]).
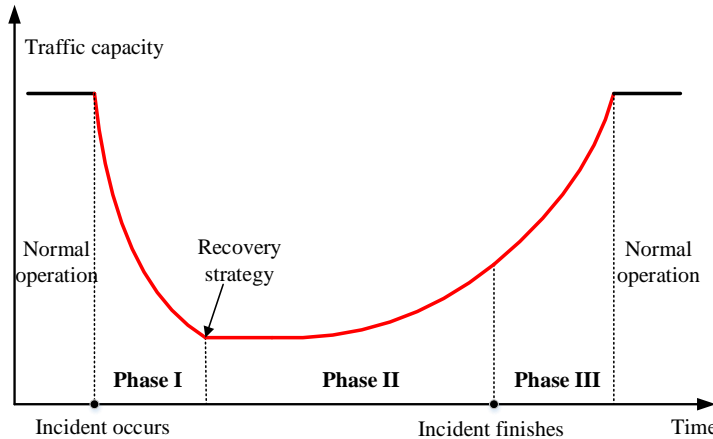

Figure 1: Three-phase disruption management in metro systems

All three phases are usually carried out by human dispatchers, and the impact of the recovery actions depends critically on the level of expertise of the people in charge. Additionally, default rescheduling strategies focus on re-establishing the normal train traffic flow, with little or no concern for passengers' inconvenience. In contrast, since the passengers are the actual target of the offered service, every recovery action should also aim at improving the quality of the service offered by explicitly taking into account the estimated passenger flow.

The purpose of this paper was to develop models and algorithms to support dispatchers in all phases of the disruption management. To this end, we need to address both the off-line problem of choosing the recovery action and the online problem of controlling trains in real-time. In the off-line phase we limit our developments to two common recovery strategies for ring lines[1]. The first strategy consists of alternating

---

[1]We focus on this type of network topology because our real-life test case, Line 2 in Beijing underground system, is a ring line. Also, other, less frequent recovery actions could be addressed. Indeed, extending our models and methods would be rather straightforward, but is out of the scope of the current paper.

trains in both directions on the single track parallel to the failure track (called SMALLCAPS ALTERNATE), whereas the second consists of short-cutting the double ring structure, culminating with two double-ring sublines that share a short section where they exchange passengers (called SMALLCAPS SHORTTURN). Correspondingly, two nonlinear mixed integer programming (NMIP) models are firstly formulated to cope with the complicated interaction between train scheduling and passenger flow management. To satisfy real-time computing requirements, we propose a novel method coupling time-index and big-M formulations to linearize NMIP model as the mixed integer linear programming (MILP) model. Then, for the real-time information under the disruption scenario (like passenger demands and the duration time), a two-stage approach is also designed by solving the proposed MILP models: in Task 1, executed in Phase 1, the best recovery strategy is selected with a series of historical data to obtain the initial timetable; in Task 2, executed in Phase 2, the initial timetable is iteratively updated for the real-time rescheduling according to the rolling optimization.

Therefore, during Task 1, our method identifies the best recovery action. Previous works discuss several strategies, such as alternate operation [37], holding train [9, 25], deadheading [11], station skipping [12], short-turning with a fixed crossover track [29], but only focus on one of these targets, i.e., passenger or train movement recoveries. In Task 2, the optimal strategy is implemented. Because the overall train traffic flow is shaken up, the selected optimization model is now exploited at the operational level to schedule in real-time movements of trains along the line according to the selected recovery strategy. The objective is the minimization of a combination of passenger inconvenience and deviation from the wanted timetable. This is the online (or *closed-loop*) phase of our approach, also tackled in [33] and [19]. The online re-optimization thus allows better consideration of input data (as past passenger flows become known). However, there is also an additional and crucial benefit: indeed, the duration of the disruption is often underestimated, and the initial prediction can be corrected when better estimations are at hand.

Train scheduling problems, which include the (online) train dispatching and the (offline) train timetabling problem, fall into the class of job shop scheduling problems ([28, 23]). There are basically two MILP models for this class of disjunctive problems: big-M formulation and time-indexed formulation (a third approach has been recently introduced in [17, 22]). For a full introduction and comparisons of these two formulations (on single machine scheduling problems) we refer the reader to [30] (theoretical discussions) and [15] (computational tests). As discussed in [30, 15], big-M formulations return poor bounds and consequently generate large search trees when embedded in branch & bound schemes. However, linear relaxations are solved very efficiently in short computing time. In contrast, time-indexed formulations (see [4, 6, 39]) are much tighter because the number of binary variables grows linearly with the number of periods in a discretized time horizon, and the relaxations are much harder to compute. The computational experience for single machine scheduling problems in [15] shows that the trade-off is in favor of the big-M formulations. This trade-off is even more evident in the experience on real-time train rescheduling in terminal metro stations reported in [21], where the time period in the discretization was fixed to 5 seconds by request of the subway engineers.

Nevertheless, in train scheduling, some works adopt time-indexed formulations with different solution techniques (see [5, 24, 3, 38, 13]), especially for offline problems, where one can exploit longer computing times. However, for the reasons described, big-M formulations are normally the preferred option (for instance, [10, 8, 16, 32, 31]) for online train rescheduling problems. This option was also our initial call. Informally, in our model, we need to synchronize two different submodels: one for train operations, and the other for passenger flows. However, our attempts to extend in a natural way the classical big-M model developed for train operations by linearizing passenger flows have failed because arrival rates vary

over time (fixed arrival rates are considered in [40]). Additionally, the alternative approach, namely, a full time-indexed formulation, turned out to be unfeasible. Indeed, if we choose large discretization steps, the solutions produced are unrealistic. With smaller discretization steps, the resulting MILP becomes too large to be solved by state-of-the-art solvers. Interestingly, in the recent paper [35], which also combines passenger flows and train scheduling, a full time-indexed formulation is successfully applied to compute solutions by employing a smart discretization mechanism. We decide to investigate another direction.

In particular, we develop a novel MILP model, which embeds two major blocks: a "big-M block" for modeling train movements, and a "time-indexed block" to model passenger flows over time. The key observation is that a 5 minute-period discretization appears to be a reasonable approximation to represent passenger flow dynamics so that the size of the time-indexed block can be kept at bay. By exploiting this idea, we finally manage to tackle our experimental instances in the available computing time. Note that the model makes use of passenger arrival rates at each platform and each point in time, which are uncertain parameters. The sensitivity of timetables with respect to noise and disturbances has been the subject of several studies (see, for instance, [7, 18]). In our case, we can take advantage of the availability of many historical samples of passenger arrival rates. Rather than selecting (or calculating) a particular sample matrix, we extend our basic MILP model to evaluate our objectives with respect to a set of arrival rate matrices.

In summary, the main contributions of this paper to the current body of research are the following:

- Modeling and solving the decision problem associated with the offline recovery strategy selection and the online implementation of the best adaptive schedule.

- Including the passenger satisfaction and headway deviation as main components in the objective function and, to this end, explicitly handling passenger flows with arrival rates that vary over time.

- Developing a novel formulation by combining big-M and time-indexed models.

- Testing our approach over real-life instances from Line 2 of the Beijing metro system.

For comparative convenience, the detailed characteristics of some closely related references are summarised in Table 1.

Table 1: Characteristics comparison of some closely related studies

Publication	Objective[1]	Recovery strategy (online)[2]	Linear formulation[3]	Solution algorithm
Ghaemi et al. [14]	TD	Short-turn (no)	Big-M	Gurobi
Xu et al. [37]	TD	Alternate operation (no)	Big-M	Heuristic algorithm
Mannino and Mascis [21]	TD	Interlocking route of terminals (yes)	Big-M	Branch and bound
Sánchez-Martínez et al. [33]	PTT	Holding train (yes)	-	Non-linear optimizer
Gao et al. [12]	PTT, NBP	Station skipping (no)	-	Heuristic algorithm
Yin et al. [38]	PTT, TEC	Running time (no)	Time-indexed	Approximate dynamic programming approach
This paper	NBP, HD	Alternate and short-turn operations (yes)	Coupling	Two-task approach with CPLEX

[1] Symbols description of objective: train delay time or cost (TD), passenger travel time (PTT), train energy consumption (TEC), number of nonboarding passengers (NBP), headway deviation (HD).
[2] Online means whether the reference is applied to online scheduling with real-time dynamic information.
[3] Linear formulation includes big-M and time-indexed formulations. "Coupling" means coupling those two formulations.

The rest of this paper is organized as follows. The problem is introduced in Section 2. In Section 3, we describe the proposed MILP models, and in Section 4, we describe the overall two-stage approach to handle the rescheduling problem from both offline and online perspectives. Our experiments with real-life

data from Beijing metro Line 2 are implemented in Section 5, and final conclusions are given in Section 6.

2   Problem description

We consider a double-track, bidirectional double-loop line as in Figure 2, including N stations, $2N$ stopping platforms (1 platform per station and per direction) and some crossover tracks. The set of platforms is denoted by $\mathcal{U} = \{1, \ldots, 2N\}$. During normal operations, the inner loop is traveled clockwise, through platforms $1, \ldots, N$. The outer loop is run in the opposite direction through platforms $N+1, \ldots, 2N$. We have two platforms at each station, one in the inner loop and the other in the outer loop, with the matched platforms $(1, 2N), (2, 2N-1), \cdots, (N, N+1)$. Moreover, the terminal platforms $1, N, N+1, 2N$ are connected to the external depot, where trains enter and exit the line. In normal operations, an inner train leaves the depot to start its first cyclic service from platform 1; then, it runs through the sequence[2] $(1, 2, \ldots, N, 1)$ of platforms and then starts the next service at platform 1. Similarly, in the outer loop, each cyclic service will go through the sequence $(N+1, N+2, \ldots, 2N, N+1)$. The inner and outer loops are connected by crossover tracks. Crossovers can only be used in emergency situations.
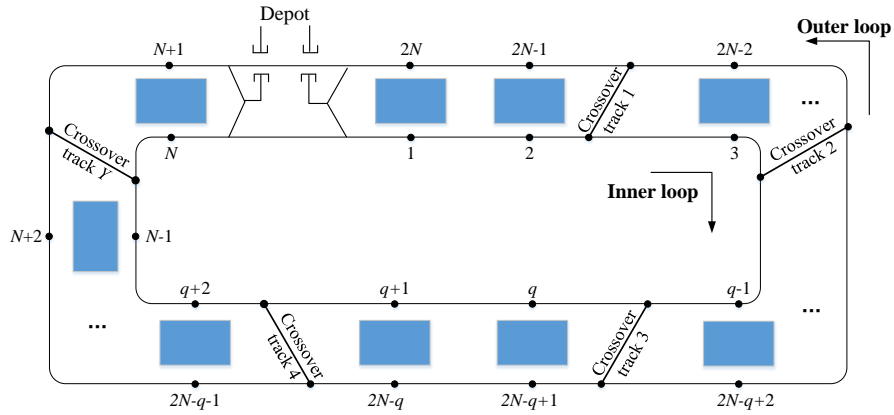

Figure 2: The structure of the rail transit line

We select the single blockage of one inter-station caused by a power failure (other reasons are discussed in Section 1). Normally, each traction substation transforms the voltage from high to low and provides for adjacent sections of upstream and downstream through the overhead line/third rail. Due to electricity tripping, power supply failure of one direction will affect the provided voltage of the other direction. Hence, we assume that the corresponding section of track blockage will be limited to only one train operation. Note that, this assumption could be modified for different types of disruptions.

In the example of Figure 3, we depict a track blockage in the inner loop that is in the track section between platforms q and $q + 1$, for instance, because of a power failure. In this track section, trains without onboard storage devices cannot operate until the failure is repaired. In addition, for power supply reasons, we assume that only one train at a time is allowed to operate on the corresponding outer section, namely, the track section between platforms $2N - q$ and $2N - q + 1$ (which are matched with inner platforms $q + 1$ and q, respectively).

---

[2]Train routes (or subroutes) will be denoted by the ordered list of platforms $(s_1, s_2, \ldots)$. The track section between two successive platforms, for example, u and w, is denoted by $[u, w]$.

In Figure 3, we have 5 service trains operating on the transit line, and 14 groups of passengers $(P1, \ldots, P14)$ waiting at different platforms: $P1, \ldots, P7$ are in the inner loop, whereas $P8, \ldots, P14$ are in the outer loop. Trains 1, 4 and 5 are in the outer loop with Train 1 in section $[2N-2, 2N-1]$, Train 4 in $[2N-q, 2N-q+1]$ and Train 5 in $[2N-q-2, 2N-q-1]$. For the inner loop trains, Train 2 is in section $[1, 2]$ and Train 3 in section $[q-1, q]$. When the disruption occurs, nearby Trains 3, 4 and 5 will be held at the next platforms where they wait for new instructions until a recovery timetable becomes available. As a consequence, passengers $P1, \ldots, P14$ queue up at platforms q, $q+1$, $2N-q$ and $2N-q+1$, and wait for the service recovery.
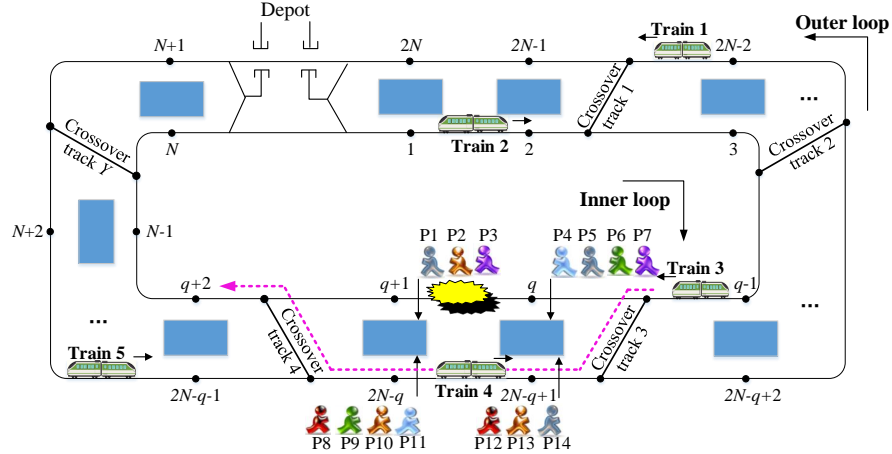
(a) ALTERNATE strategy
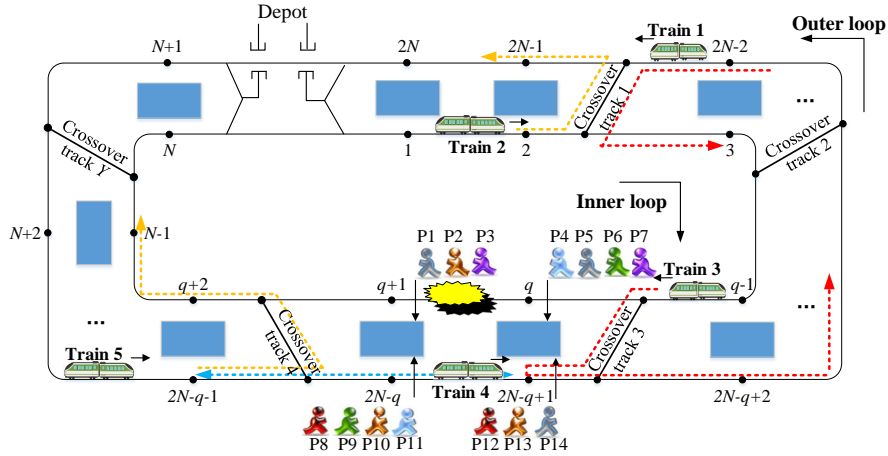
(b) SHORTTURN strategy

Figure 3: Two train recovery strategies

2.1 Recovery strategies

In this paper, we compare two recovery strategies when one track fails: the first with *alternate* operations, and the second with *short-turn* and bidirectional operations. Both strategies exploit the availability of crossovers. Figure 4 may help understand the role of the crossovers in both strategies.

In the figure, the disruption occurs in the track between platforms E_D and S_D in the inner loop (later denoted as *bad loop* because it contains the disrupted track, whereas the other loop is the *good loop*). Traversing the bad loop according to the natural orientation - namely, the one followed by trains during
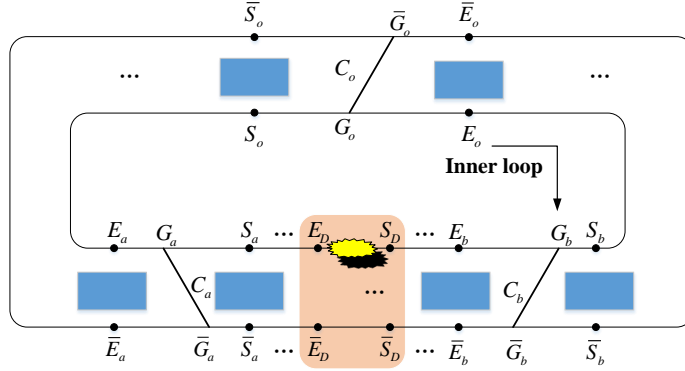
Figure 4: The two "near-by" crossovers C_a and C_b, and the "opposite" crossover C_o

normal operations, we denote by C_a the first crossover encountered on the loop after the failure track, and by C_b the last crossover encountered before the failure track (see Figure 4). Crossover C_a meets the inner loop in a point G_a and the outer loop in point $\bar{G}_a$; crossover C_b meets the inner loop in point G_b and the outer loop in point $\bar{G}_b$. Additionally, the first platform on the inner loop after G_a (G_b) is E_a (E_b), corresponding to platform $\bar{E}_a$ ($\bar{E}_b$) on the good loop.

ALTERNATE strategy. In this setting, trains are allowed to crossover between the inner and the outer loops in proximity of the failure track during recovery phases. Then, the bad-loop trains are redirected to the good loop at the first crossover C_b and then return to their original loop at the second crossover C_a. In this way, the failure track is skipped. The tracks on the good loop between C_b and C_a (precisely, between points $[\bar{G}_a, \bar{G}_b]$ in the figure), which are parallel to the skipped tracks on the bad loop, will be traversed by trains in both directions. More specifically, the trains on the bad loop will take the detour $S_b \to G_b \to \bar{G}_b \to \bar{G}_a \to G_a \to E_a$.

To clarify this scenario, we observe the example of Figure 3(a). Here, C_a and C_b correspond to Crossover 4 and Crossover 3, respectively. Train 3 cannot continue on its normal route because of the disruption of section $[q, q+1]$. Instead, it travels through Crossover 3, continues on the section $[2N-q+1, 2N-q]$ (in the opposite direction) and then takes Crossover 4 to regain the normal route. The modified subroute $(q-1, 2N-q+1, 2N-q, q+2)$, is represented by the pink dotted line. Accordingly, the inner passengers will move from the inner platform to the matched outer platform: namely, $P1, \ldots, P3$ will move to $2N-q$, whereas passengers $P4, \ldots, P7$ move to $2N-q+1$ and then wait for the next inner trains. The outer passengers $P8, \ldots, P14$ will stick to their plans and board the outer Train 4 at the original platforms. With this strategy, trains will alternate in both directions on track section $[2N-q, 2N-q+1]$ with a drastic reduction of frequency and thus of capacity. Especially in peak-hours, not all passengers will be able to board the first incoming train, and the average number of waiting and nonboarding passengers will increase (along with their dissatisfaction). This increase is particularly true for platforms at an increasing distance from the disrupted track along the inner route because trains may easily reach their capacity at earlier platforms. Because of the reduced capacity over the single-track region between the two crossovers, this strategy cannot be applied when the number of platforms on the single-track region is too large.

SHORTTURN strategy. In this strategy, we make use of a shuttle train in the segment parallel to the failure track in order to guarantee the correct headway between successive trains on the rest of the

line. In addition to the two crossovers C_b and C_a (before and after the failure track) introduced for the ALTERNATE strategy, we pick a third, "opposite" crossover C_o as follows. Note first that C_b and C_a split each of the original loops into two segments. For the bad loop, one segment contains the failure track, and we call the remaining part the *opposite segment*. Similarly, the good loop is also split into two segments, one "parallel" to the failure track, and an opposite segment. Crossover C_o is chosen to connect the opposite segment in the bad loop to the opposite segment in the good loop. If no such crossover exists, the strategy cannot be applied. Note that, the SHORTTURN strategy has two advantages: (1) A much flexible operation mode can be taken into consideration. By optimizing the recovery strategy, the best location of crossover track C_o can be found through considering the pre-specified/historical passenger demands. The optimal recovery strategy can expectedly improve the system efficiency as much as possible. (2) The SHORTTURN strategy is helpful for transporting the short-distance passengers inside the left-/right-loop, since the train needs to unload all passengers before it turns around on the crossover tracks, which in essence provides more space for passengers and improve the transportation efficiency. Especially, we note that the short-turn can also occur on the disrupted section in a special case, when C_o overlaps with the crossover track C_a or C_b.

In Figure 4, three crossovers meet the inner loop in three points, i.e., G_b, G_a, G_o (in clockwise order), with the failure track between G_b and G_a. Similarly, three corresponding points where the crossovers meet the outer loop will be $\bar{G}_b, \bar{G}_a, \bar{G}_o$ (in clockwise order). We can then identify two new loops. "Right" loop b, is obtained by concatenating Crossover C_b, the section of outer loop from $\bar{G}_b$ to $\bar{G}_o$, Crossover C_o, the section of inner loop from G_o to G_b, and platform $\bar{E}_b$. In addition, "left" loop a, is obtained by concatenating Crossover C_a, the section of outer loop from $\bar{G}_a$ to $\bar{G}_o$, Crossover C_o, and the section of inner loop from G_o to G_a. Trains will be rerouted on these two new loops. In particular, some trains will run on Loop b and other trains on Loop a. The passengers between $\bar{G}_a$ and $\bar{G}_b$ will be served by a bidirectional shuttle. The bidirectional shuttles will "penetrate" the left loop a to the first station $\bar{E}_a$, where it will exchange passengers with the left loop trains. On the other side, things go slightly different, as the right loop trains are now "penetrating" the bidirectional section until the first station $\bar{E}_b$ to the left of $\bar{G}_b$, in order to exchange passengers with the shuttles.

Letting C_b, C_a and C_o correspond to Crossover 3, Crossover 4 and Crossover 1, respectively, this strategy is depicted in Figure 3(b). Here, the line is divided into a left loop (between Crossover 1 and Crossover 4), a bidirectional track (between Crossover 3 and Crossover 4) and a right loop (between Crossover 1 and Crossover 3), where the corresponding operation routes are shown in orange, blue and red dotted lines, respectively. To connect the left and right loop passenger flows, a bidirectional shuttle operates between the outer platforms $2N - q - 1$ and $2N - q + 1$. The choice of the suitable boundary crossover C_o depends on the expected number of passengers at the platforms and train capacity.

In the example of Figure 3(b), Trains 1 and 3 start the short-turn operations on the right loop, and Train 4 begins the bidirectional operation from platform $2N - q + 1$. On the left loop, Trains 2 and 5 operate the service for the left loop. Passengers $P1, \ldots, P3$ and $P8, \ldots, P11$ can only take Train 4 at platform $2N - q$ to $2N - q - 1$ or $2N - q + 1$ to transfer to left-loop or right-loop trains to their destinations. When the destination platform belongs to the right loop, passengers can directly take the right-loop Train 3 at platform $2N - q + 1$ after alighting the bidirectional Train 4; otherwise, the passenger should change the platform from $2N - q - 1$ to $q + 2$ to transfer to left-loop trains.

2.2 Passenger arrival process

Passengers arrive over time at stations where they wait, board, alight and, possibly, transfer from one train to another. In our case, transferring is only considered in the SHORTTURN strategy. In principle, the arrival rate of new passengers at stations could be immediately derived by time-dependent origin-destination (OD) matrices (see, e.g., [38], [1], [24]). However, OD matrices are seldom available. Nevertheless, arrival rates can be easily collected in real time at each station.

We assume we have at hand for every platform u the (expected/historical) passenger arrival rate over time, which is a continuous function $\bar{p}^u(t)$ of the time t. Observe that if $\Delta \in \mathbb{R}$ is sufficiently small, we may assume that $\bar{p}^u(t + \Delta) \approx \bar{p}^u(t)$. This assumption allows us to sample the function in a discrete set of points. In particular, we discretize the planning horizon $[T_{sta}, T_{end}]$ (namely, Phase II in Figure 2) into $Q - 1$ periods of equal size δ, starting at $T_{sta} = T_1, T_2, \ldots, T_{Q-1}$, respectively, and with $T_{end} = T_Q$. Then, the arrival rate p_i^u is defined as the number of passengers arriving at platform $u \in \{1, \ldots, 2N\}$ in period $i \in \{1, 2, \ldots, Q - 1\}$. We let $P = [p_i^u]$, $i = 1, \ldots, Q - 1$, $u \in \mathcal{U}$ be the arrival rates matrix.

In our model, we will make use of a historical set $\mathcal{P}^0 = \{P^1, \ldots, P^{RS}\}$ of arrival rates matrices that are sampled over a suitable period of time, where, for $r = 1, \ldots, RS$, we have $P^r = [p_i^{u,r}]$, $i = 1, \ldots, Q - 1$, $u \in \mathcal{U}$.

To better understand the passenger traveling process, consider the example illustrated in Figure 5. The black lines are the planned timetables, whereas the red lines are the rescheduled timetable, after a disruption has occurred. Bidirectional Train l and right-zone Trains $k - 1$ and k, respectively, stop at inner platform q and the relevant outer platform $2N - q + 1$. In the figure, we also show 17 groups P1,..., P17 (50 passengers per group) waiting at platforms q and $2N - q + 1$, respectively, including transferring passenger groups P1, P12 and P13. We assume that the average boarding rate is 1 person/s and the maximum dwell time is 300 seconds. In this illustration, each train can load up to 500 passengers. The
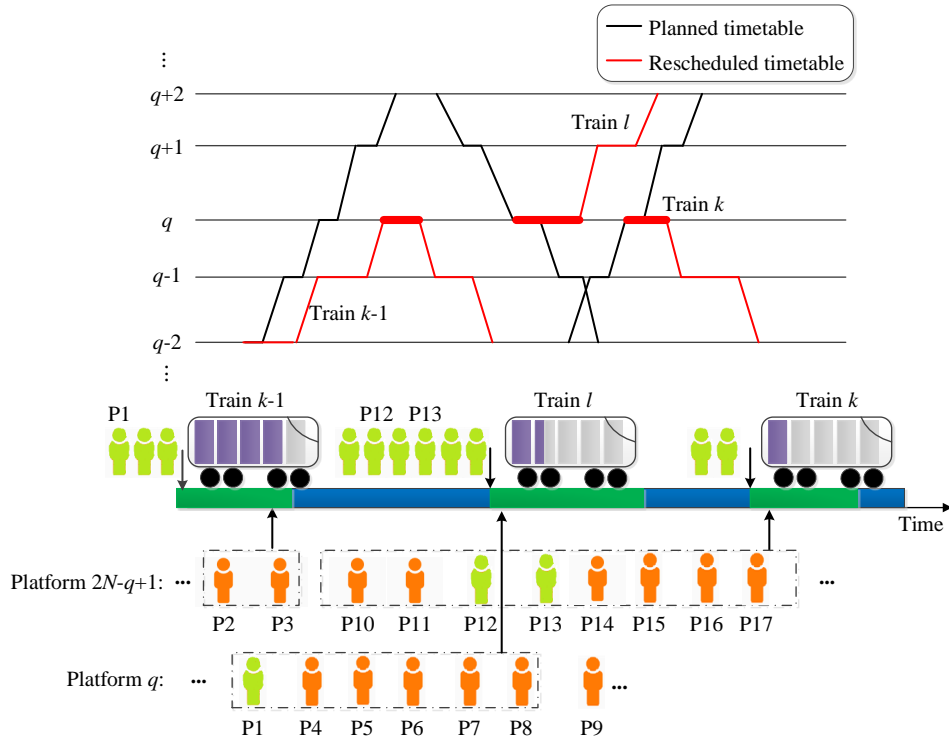

Figure 5: Passengers waiting, boarding, alighting and transferring at platform $2N - q + 1$

capacity is represented in each train by 5 rectangles, each associated with a load of 100 passengers. The light gray rectangles represent the residual capacity after passengers alighting, so Trains $k-1$, l and $k+1$ can load at most 100, 350 and 400 passengers, respectively, corresponding to 2, 7 and 8 groups. Thus, when Train $k-1$ arrives at platform $2N-q+1$, only two groups of waiting passengers (i.e., P2 and P3) can go on board. In principle, the next arrival Train l could accommodate up to 7 groups of passengers (from P4 to P9 and transferring P1). However, since the maximum dwell time is 300 seconds, Train l can board 300 passengers, i.e., 6 groups. Finally, Train k will accommodate groups P10,..., P17, including transferring passengers P12 and P13 from Train l.

2.3 Objectives

When choosing an appropriate rescheduling strategy, we consider the trade-off between regaining the regular passenger transport capacity and train regularity. In particular, we consider a term for the number of passengers unable to board on the first incoming train and a term for the average headway deviation in the objective. Observe that maintaining the desired train headway is crucial for trains quick recovery to the planned timetable.

Assuming the fixed capacity of trains and fleet sizes, the number of transported passengers is essentially associated with dwell times, train frequency, train residual capacity and the actual number of passengers waiting at platforms. Note that the two terms in the objective function are somehow competing. Indeed, longer dwell times allow more passengers to board trains but increase headway deviations.

Particularly, when the passenger demand is much less, and all the waiting passengers can board the first arrival train, the passenger flow can be neglected in the model since the second objective is essentially co-monotone with the first one. However, during the disruption, a large amount of passengers might be accumulated on each platform, and the passenger demand could exceed the total train capacity, above all, in the peak hours. In this case, these two objectives are not necessarily co-monotone due to different distributions of the passengers on individual platforms. Thus, it is necessary for us to take passenger demands into consideration specifically in order to transport the accumulated passengers as soon as possible. In our model, we will make the following assumptions.

1. Between successive platforms, there is only one track per direction. Additionally, in each station there is only one platform per direction. Crossovers can accommodate at most one train at a time. Trains can only stop at platforms (i.e., not on the tracks between platforms), and they are not allowed to skip the stop at platforms.

2. During the disruption period, the fleet size does not change, and the running time on each track section and each crossover is a fixed, known parameter.

3. No waiting passengers are leaving the station without boarding (to adopt a different traffic mode).

4. No more than one train at a time can occupy the section on the good loop parallel to the failure track for the power supply failure (in the example of Figure 3(b), the section between platforms $2N-q$ and $2N-q+1$).

3 Mathematical model

In this section, we develop MILP formulations for the rescheduling problems associated with the two strategies in the previous section. We assume passengers arrive at platforms over time with a given

projected rate. The fleet of trains in each direction has fixed size V. The platforms in the inner loop will be labeled as $1, \ldots, N$, while in the outer loop as $N + 1, \ldots, 2N$. We introduce here some notation to denote the position of the three relevant crossovers (see Figure 4):

- The blockage occurs in the inner loop in the section $[S_D, E_D]$, parallel to section $[\bar{E}_D, \bar{S}_D]$ on the outer loop.

- C_b meets the inner loop at point G_b on the track $[S_b, E_b]$, and the outer loop at point $\bar{G}_b$ on the parallel track $[\bar{E}_b, \bar{S}_b]$;

- C_a meets the inner loop at point G_a on track $[S_a, E_a]$, and the outer loop at point $\bar{G}_a$ on track $[\bar{E}_a, \bar{S}_a]$;

- C_o meets the inner loop at point G_o on track $[S_o, E_o]$, and the outer loop at point $\bar{G}_o$ on track $[\bar{E}_o, \bar{S}_o]$ [3].

Extending our model to a more general setting is straightforward but would require more complex definitions without adding particular insight.

We start now by describing parameters, variables, and constraints associated with train operations. Then, we will focus on modeling passenger flows and finally, we discuss the objective function. Note that in normal operations, the number K of trains in each direction is constant. During the disruption, the total number of trains ($2K$) does not change; however, the number of trains running through the different sections in which the line is decomposed will depend on the adopted strategy, in particular, on the position of the boundary crossover C_o (see Figure 4) and on the time when the strategy is implemented. For clarity, the related parameters and variables are summarized in Table 2.

Table 2: Related parameters and variables in the formulation

Symbol	Definition
$[T_{sta}, T_{end}]$	disruption duration interval
$\mathcal{U}$	set of platforms, $\mathcal{U} = \{1, \ldots, 2N\}$
$\mathcal{K}$	set of service trains, $\mathcal{K} = \{1, \ldots, 2K\}$
$\mathcal{Q}$	set of periods in planning time horizon, $\mathcal{Q} = \{1, \ldots, Q\}$
$\mathcal{R}$	set of passenger arrival rate samples, $\mathcal{R} = \{1, \ldots, RS\}$
$\mathcal{K}_I / \mathcal{K}_O$	inner/outer loop trains (ALTERNATE strategy)
$\mathcal{K}_L / \mathcal{K}_R / \mathcal{K}_B$	left/right/bidirectional trains (SHORTTURN strategy)
$R^{u, v(u)}$	running time from platform u to next platform $v(u)$
Θ_c	required time for switch transaction of crossover track
s_{min} / s_{max}	minimal/maximal dwell time at any platform
H_{min}	minimal headway between successive trains
H_p	planned headway between two successive service trains
R_b	boarding passenger ratio at each time
C	loading passenger capacity of single service train
A^u	alighting rate of in-vehicle remaining passengers at platform u, $A^u \in [0, 1]$
d_k^u	departure time of service train k at platform u
a_k^u	arrival time of service train k at platform u
s_k^u	dwell time of service train k at platform u
y^{kl}	binary variable that is 1 if train k precedes l in the shared region; otherwise, 0
p_i^u	passenger arrival rate at platform u at time period i, $i \in \mathcal{Q} \setminus \{Q\}$
pw_k^u	number of passengers waiting for train k at platform u
pb_k^u	number of passengers boarding on train k at platform u
pa_k^u	number of passengers alighting from train k at platform u
pr_k^u	number of passengers remaining in train k after departing from platform u

---

[3] We remark here that the position of crossover C_o is indeed one of the decisions to take in Phase 1 with the help of our model. However, to simplify the following discussion and notation, here we consider for C_o only one position.

3.1 Modeling train operations

For train $k \in \mathcal{K}$, we denote by d_k^u, a_k^u, $s_k^u \in \mathbb{R}_+$ the (variable) arrival, departure and dwell time, respectively, at platform $u \in \mathcal{U}$. In the normal operation, we let $\mathcal{K}_I = \{1, \ldots, K\}$ be the set of inner loop trains and $\mathcal{K}_O = \{K+1, \ldots 2K\}$ be the set of outer loop trains. Therefore, the first train in the inner direction will be train 1, whereas the first train in the outer direction will be $K+1$. With the ALTERNATE strategy, the number of trains in each direction does not change w.r.t. the normal regime; therefore, we keep the same notation (even if now the outer and inner loops have a short overlap). For the SHORTTURN, the set of all trains $\mathcal{K}$ is partitioned into three sets $\mathcal{K}_L, \mathcal{K}_R, \mathcal{K}_B$, i.e., the sets of left loop trains, of right loop trains, and of bidirectional trains, respectively. The exact size of each set, in this case, depends on the specific moment in which Phase II begins and the position of the shortcut. In any case, all the sets of trains are naturally ordered. We also assume that trains are ordered, and with some abuse of notation, for $k \in \mathcal{K}_I$ ($k \in \mathcal{K}_O$, $k \in \mathcal{K}_L$, $k \in \mathcal{K}_R$, $k \in \mathcal{K}_B$) we denote by $k+1$ the train following k in $\mathcal{K}_I$ ($\mathcal{K}_O, \mathcal{K}_L, \mathcal{K}_R, \mathcal{K}_B$). Finally, observe that several constraints are common for both strategies. For ease of notation, we only write such constraints for the ALTERNATE strategy [4].

3.1.1 Timetable constraint

(1) Running time constraint Denoting by $R^{u,v(u)}$ the time necessary for a train to run from platform u to platform $v(u)$, we have

$$a_k^{v(u)} - d_k^u = R^{u,v(u)}, \quad u \in \mathcal{U}, k \in \mathcal{K} \tag{1}$$

Note that in normal operations $v(u) = u+1$ for $u \neq N$ and $u \neq 2N$; if $u = N$ ($u = 2N$) then the train will start a new cycle in $v(u) = 1$ ($v(u) = N$). In this case, $R^{u,v(u)}$ also includes the additional dwell time required for the change of train service. Finally, when trains cross over between the loops, u and $v(u)$ assume values that depend on the crossing points, and the constant $R^{u,v(u)}$ will clearly also embed the time to run the crossover.

(2) Dwell time constraint Next, the dwell time of a train at a platform is limited in a given range:

$$s_{min} \leq d_k^u - a_k^u \leq s_{max}, \quad u \in \mathcal{U}, k \in \mathcal{K} \tag{2}$$

where s_{min}, s_{max} are input parameters that are equal for all trains.

(3) Headway constraint Consecutive trains must hold a safe distance. Assuming a constant speed on all tracks, this distance becomes[5]:

$$d_k^u - d_{k-1}^u \geq H_{min}, \quad u \in \mathcal{U}, k \in \mathcal{K} \setminus \{1, K+1\} \tag{3}$$

where H_{min} is the minimum required headway.

---

[4]For the similar reason, in the rest parts of this section, we also apply ALTERNATE strategy to model passenger flow and objective.

[5]Actually, H_{min} may also depend on the platform u, but we neglect this here for sake of brevity. Also, this constraint can easily take into account headway between trains at platforms.

3.1.2 Disruption dependent constraints

Recall that we can have only one train at a time on the bidirectional section $[\bar{E}_D, \bar{S}_D]$. Consequently, a train cannot occupy the track before the leading train (in the same direction) has cleared it. For a pair of consecutive trains $k-1$ and k, this condition is translated into

$$d_k^{\bar{E}_D} - a_{k-1}^{\bar{S}_D} \geq 0, \quad k \in \mathcal{K}_O \setminus \{K+1\}$$
$$d_k^{\bar{S}_D} - a_{k-1}^{\bar{E}_D} \geq 0, \quad k \in \mathcal{K}_I \setminus \{1\}$$

(4)

where the different orderings of the platforms depend on the direction of the trains.

3.1.3 Sequencing constraints

When applying recovering strategies, trains from the two loops end up running through the same line resources. When this happens, we need to sequence them correctly on the shared tracks and crossovers.

In the ALTERNATE strategy (described in Figure 4), trains in both directions share the outer section $[\bar{S}_a, \bar{E}_b]$, but the section can accommodate only one-way trains at a time. For any pair of trains in opposite directions, for example, $k \in \mathcal{K}_\mathcal{I}$ and $l \in \mathcal{K}_O$, we introduce a binary variable y^{kl} that is 1 if k precedes l on the section; otherwise, 0. Then, we have:

$$d_k^{S_b} - a_l^{\bar{S}_b} \geq \Theta_c - My^{kl},$$
$$d_l^{\bar{E}_a} - a_k^{E_a} \geq \Theta_c + M(y^{kl} - 1), \quad k \in \mathcal{K}_\mathcal{I}, l \in \mathcal{K}_O$$

(5)

where Θ_c is the time needed to switch a crossover track, and M is a suitably large constant. We here use the classical big-M trick: indeed, when $y^{kl} = 1$, the first constraint is always satisfied (by otherwise feasible departures and arrivals), and thus it becomes redundant. Conversely, when $y^{kl} = 0$, the second constraint becomes redundant.

Remark 1. *In the* SHORTTURN *strategy (see Figure 4), the line is divided into three parts by crossovers tracks C_b, C_a and C_o. Trains running on the right loop and on the left loop will share the unit capacity region around crossover C_o, and they need to alternate on it. Therefore, for $l \in \mathcal{K}_L$ and $k \in \mathcal{K}_R$, we introduce the binary variable y^{lk} that is 1 if l precedes k in the region; otherwise, 0. Then, we have:*

$$d_l^{S_o} - a_k^{E_o} \geq \Theta_c - My^{lk},$$
$$d_k^{\bar{E}_o} - a_l^{\bar{S}_o} \geq \Theta_c + M(y^{lk} - 1), \quad l \in \mathcal{K}_L, k \in \mathcal{K}_R$$

(6)

Trains running on the left loop and the bidirectional trains in the shared outer section $[\bar{E}_a, \bar{S}_a]$ also need to alternate. For $l \in \mathcal{K}_L$ and $j \in \mathcal{K}_B$, binary variable y^{jl} is 1 if j precedes l in the shared region; otherwise, 0. Additionally, because a bidirectional train $j \in \mathcal{K}_B$ goes through intermediate platforms both in the left and right directions, when necessary, we duplicate arrival and departure variables. Therefore, we have

$$d_l^{\bar{E}_a} - \overrightarrow{d_j^{\bar{E}_a}} \geq H_{min} + M(y^{jl} - 1),$$

$$\overleftarrow{d_j^{\bar{S}_a}} - a_l^{E_a} \geq \Theta_c - My^{jl}, \qquad\qquad l \in \mathcal{K}_L, j \in \mathcal{K}_B \qquad\qquad (7)$$

$$d_l^{\bar{E}_a} - \overrightarrow{a_j^{\bar{S}_a}} \geq \Theta_c + M(y^{jl} - 1),$$

The first constraint ensures the minimum headway H_{min} when the bidirectional train j precedes the left loop train l. Observe that since train l can only leave platform $\bar{E}_a$ moving towards the right, we do need to duplicate the variable. The second and third constraints in (7) allow only one train at a time in the shared region. Here, $\overrightarrow{d_j^{\bar{S}_a}}$ denotes a departure towards the right, whereas $\overleftarrow{a_j^{\bar{S}_a}}$ is an arrival when moving leftwards to the terminal $\bar{E}_a$.

When considering the right loop trains, first observe that in our strategy they overlap with bidirectional trains only in the first platform $\bar{E}_b$ (see Figure 4). The sequencing process can thus be modeled similarly to the last two constraints in (7) after introducing suitable variables and constants (which we omit writing for the sake of brevity).

3.2   Modeling passenger flow

As described in Section 2, at platform u, a number of passengers will wait for the next train k. This number depends on the passengers already waiting for the previous train $k-1$ but unable to board, on the passengers arriving at the station in the meantime, and for the SHORTTURN strategy, on transferring passengers. The number of boarding passengers depends on the spare capacity of train k, which in turn depends on the number of passengers on board when arriving at u and the number of passengers alighting at u. To describe this dynamic, we introduce several new variable quantities and some constants. For platform u and train k, pr_k^u represents the passengers on board train k when the train leaves platform u (and thus when the train arrives at the next platform); pa_k^u is the number of passengers alighting the train, pw_k^u is the number of passengers waiting when k arrives, pb_k^u is the number of boarding passengers and finally, pc_k^u is the residual capacity of the train.

Alighting passengers. We assume that the number of passengers at platform u is proportional (by a factor A^u) to the number of passengers on board when the train arrives at the platform u

$$pa_k^u = \begin{cases} pr_k^{u-1} \cdot A^u, & u \in \mathcal{U} \setminus \{1, N+1\}, k \in \mathcal{K} \\ 0, & u \in \{1, N+1\}, k \in \mathcal{K} \end{cases} \qquad (8)$$

Waiting passengers[6]**.** The passengers pw_k^u waiting for train k at platform u consist of the passengers pw_{k-1}^u waiting for the previous train which could not be boarded, and passengers arriving after the previous train has left, namely, the $p(d_{k-1}^u, d_k^u, u)$ passengers arriving at platform u in the time interval $[d_{k-1}^u, d_k^u]$. Therefore, if we denote by pb_k^u the passengers boarding on train k at u, we have

$$pw_k^u = pw_{k-1}^u - pb_{k-1}^u + p(d_{k-1}^u, d_k^u, u), \quad u \in \mathcal{U}, k \in \mathcal{K} \setminus \{1, K+1\} \qquad (9)$$

Observe that the above expression is nonlinear, as the factor $p(d_{k-1}^u, d_k^u, u)$ is a function of the variable quantities d_{k-1}^u and d_k^u. We will see in Section 3.4 how to linearize this term.

---

[6]The formulas presented here must be slightly amended when considering the transferring stations in SHORTTURN strategy. The modeling is simple but involved, requiring the introduction of auxiliary variables and constraints. For the sake of clarity and brevity, we omit here the details, which would not add much to the discussion.

Residual capacity. The residual capacity of train k at platform u depends on the empty train capacity C, the number pr_k^{u-1} of passengers in the train when the train leaves the previous platform, and the number pa_k^u of passengers alighting train at platform u.

$$pc_k^u = \begin{cases} C - pr_k^{u-1} + pa_k^u, & u \in \mathcal{U} \setminus \{1, N+1\}, k \in \mathcal{K} \\ C, & u \in \{1, N+1\}, k \in \mathcal{K} \end{cases} \tag{10}$$

Boarding passengers. The number of boarding passengers pb_k^u depends on the number of waiting passengers pw_k^u, on the dwell time s_k^u and the boarding passenger rate R_b, and on the residual capacity pc_k^u of the train after alighting passengers.

$$pb_k^u = \min\{R_b \cdot s_k^u, pw_k^u, pc_k^u\}, \quad u \in \mathcal{U}, k \in \mathcal{K} \tag{11}$$

Note that the above relationship can be easily linearized[7].

Onboard passengers. Finally, the number pr_k^u of passengers remaining onboard when train k departs from platform u is given as:

$$pr_k^u = \begin{cases} pr_k^{u-1} + pb_k^u - pa_k^u, & u \in \mathcal{U} \setminus \{1, N+1\}, k \in \mathcal{K} \\ pb_k^u, & u \in \{1, N+1\}, k \in \mathcal{K} \end{cases} \tag{12}$$

We conclude this subsection with an important remark.

Remark 2. *In our model, we have variables modeling train timetables, namely, vectors* $\mathbf{d}, \mathbf{a} \in I\!\!R^{\mathcal{K} \times \mathcal{U}}$, *and variables modeling passenger flows, namely, vectors* $\mathbf{pr}, \mathbf{pb}, \mathbf{pa}, \mathbf{pc} \in I\!\!R^{\mathcal{K} \times \mathcal{U}}$. *However, for fixed* $\bar{\mathbf{d}} \in I\!\!R^{\mathcal{K} \times \mathcal{U}}$, *then* $\mathbf{a}, \mathbf{pr}, \mathbf{pb}, \mathbf{pa}, \mathbf{pc}$ *are uniquely determined by solving the system of linear equations defined by Equation (1) and equations (8)~(12).*

3.3 Objective function

As mentioned, our objective function is the convex combination of two terms. The first term f_w is the accumulated number of passengers unable to board the first arriving train because of the limited residual capacity of the train[8]

$$f_w = \sum_{k \in \mathcal{K}} \sum_{u \in \mathcal{U}} (pw_k^u - pb_k^u) \tag{13}$$

The second term represents the positive deviation from the planned headway H_p between trains $k-1$ and k.

$$f_h = \sum_{\mathcal{K} \setminus \{1, K+1\}} \sum_{u \in \mathcal{U}} \max\{d_k^u - d_{k-1}^u - H_p, 0\} \tag{14}$$

Note that the expression for f_h can be easily linearized by replacing it with the sum of nonnegative variables q_k^u and introducing the constraint $q_k^u \geq d_k^u - d_{k-1}^u - H_p$, for $k \in \mathcal{K} \setminus \{1, K+1\}$, $u \in \mathcal{U}$.

---

[7]Because the objective function favors smaller values of waiting passengers (i.e., larger values of boarding passengers), the above nonlinear constraint can be replaced by the following linear constraints:

$$\left. \begin{aligned} pb_k^u &\leq R_b \cdot s_k^u, \\ pb_k^u &\leq pw_k^u, \\ pb_k^u &\leq pc_k^u, \end{aligned} \right\} \quad u \in \mathcal{U}, k \in \mathcal{K}$$

[8]The expression of f_w also includes an additional term to account for those passengers unable to board the last service train. We omit it here for brevity.

The overall objective function is then the weighted sum of the two terms, where the nonnegative weights λ_w, λ_h (with $\lambda_w + \lambda_h = 1$) are chosen by the planner according to the preferred target.

$$\min \ f = \lambda_w \cdot f_w + \lambda_h \cdot f_h \tag{15}$$

In practice, the setting of these two weights is dependent on the relative importance given by the dispatching policy and operators. Typically, facing with more arriving passengers in peak hours, the number of non-boarding passengers is set as the prior objective to increase transport ability, with a large value of weight λ_w. By contrast, in off-peak hours with fewer passengers, the second weight of headway deviation is favoured to regain regularity with the planned headway. Since the planned headway is pre-designed to make a trade-off between service quality and train operation cost (or other operational requirements), the price weight λ_h could be set as a larger value in this case.

Actually, in our rescheduling model, the minimal headway H_{min} is usually smaller than planned headway H_p (i.e., $H_{min} < H_p$). In this case, the solution space is further enlarged (that is, the planned headway H_p is only a feasible case of our proposed model). With the enlarged solution space, the optimal solution can be searched with the objective of the weighted average non-boarding passengers and headway deviation, where the rescheduled optimal headway can be finally deduced.

3.4 Coupling continuous and time-indexed variables

Here, we show how to linearize the term $p(d_{k-1}^u, d_k^u, u)$ in Equation (9), which describes the number of passengers arriving at a platform between two successive departures. A classical way to do this procedure is by using time-indexed formulations (as in [24, 35]) because they provide tight bounds and allow modeling complex relations by simple linear constraints. There is, however, a price to pay: the number of variables and constraints of the model increase dramatically (depending on the size of the time steps). Therefore, the time required to solve the overall problem, or even just its linear relaxation, may exceed the time allowed by the application. This has been, for instance, our experience with train rescheduling problems in metro systems, see [21]. The problem is that we need a rather aggressive discretization in order to represent train movements with sufficient details. In contrast, a grosser discretization suffices to represent passenger arrival dynamics. We will exploit this observation in the sequel of the paper to develop a model that makes use of both classical representations to describe different entities and dynamics.

We see now how to linearize the term $p(d_{k-1}^u, d_k^u, u)$ appearing in Equation (9). Recall that for the arrival matrix $P = [p_i^u]$ (or $P^r = [p_i^{u,r}]$ when looking at the r-th sample matrix), the arrival rate p_i^u at platform u is assumed to be constant in the period $i \in \mathcal{Q} \setminus \{Q\}$.

Modeling departure times. To this end, we introduce time-indexed variables and couple them with the continuous variables of our original formulation. For $k \in \mathcal{K}$, $u \in \mathcal{U}$ and $i \in \mathcal{Q}$, we introduce the binary variable $z_i^{k,u}$, which is 1 if and only if the train k departs from platform u in the time interval $[T_i, T_{i+1})$ (here we let $T_{Q+1} >> T_Q$)[9]. We also introduce the slack variable $x_i^{k,u}$, representing the offset of the departure time of train k from platform u with respect to T_i when the departure occurs in $[T_i, T_{i+1})$;

---

[9]For $\mathcal{Q} = \{1, \dots, Q\}$, the corresponding interval denotes as $[T_Q, T_{Q+1})$ if i is set as Q. Observe that the larger value of T_{Q+1} aims to contain departures times of all service trains on the whole line.

otherwise, the value of the variable is 0. Thus, we have:

$$d_k^u = \sum_{i=1}^{Q} T_i \cdot z_i^{k,u} + x_i^{k,u}, \quad u \in \mathcal{U}, k \in \mathcal{K} \tag{16}$$

Also, since departure d_k^u occurs in exactly one interval, we have

$$\sum_{i=1}^{Q} z_i^{k,u} = 1, \quad u \in \mathcal{U}, k \in \mathcal{K} \tag{17}$$

Finally, the offset is smaller than the period size:

$$
\begin{aligned}
x_i^{k,u} &\le (T_{i+1} - T_i - \varepsilon) z_i^{k,u}, \quad u \in \mathcal{U}, k \in \mathcal{K}, i \in \mathcal{Q} \setminus \{Q\} \\
x_i^{k,u} &\ge \varepsilon, \quad u \in \mathcal{U}, k \in \mathcal{K}, i \in Q
\end{aligned}
\tag{18}
$$

where ε is a suitably small constant.

Remark 3. *Observe that the "full" time-indexed formulation, proposed by previous studies (see [4, 6, 39]), is the special form of our novel formulation, in which trains depart from and arrive at each station over the discretized timestamps. For instance, the whole time horizon $[T_{sta}, T_{end}]$ is discretized as $T_1, T_2, \cdots, T_Q$ by a time interval δ. Then, the departure time can be expressed as Equation (19) for the "full" time-indexed formulation, in which binary variable $z_i^{k,u}$ takes 1 if train k departs from platform u at time T_i, and its value takes 0 otherwise.*

$$d_k^u = \sum_{i=1}^{Q} T_i \cdot z_i^{k,u}, \quad u \in \mathcal{U}, k \in \mathcal{K} \tag{19}$$

where

$$\sum_{i=1}^{Q} z_i^{k,u} = 1, \quad u \in \mathcal{U}, k \in \mathcal{K}$$

Comparing with these two methods, the expression of the novel formulation (see Equation (16)) is the same as the "full" time-indexed formulation if letting slack variable $x_i^{k,u}$ equals to 0. However, for accuracy, the "full" time-indexed formulation requires a much smaller time interval δ to discretize the time horizon than our proposed method, leading to more decision variables and longer computational time in the optimization process (Please see the comparison results in Section 5.6).

In Figure 6 we exemplify (part of) a feasible solution. The horizon $[T_{sta}, T_{end}]$ is discretized into $T_1, T_2, \ldots, T_Q$. Two trains k and $\tilde{k}$ depart from station u at time $d_k^u \in [T_2, T_3)$ and $d_{\tilde{k}}^u > T_Q$. Correspondingly, binary variables $z_2^{k,u}$ and $z_Q^{\tilde{k},u}$ are both 1 (while all other z-variables are 0). For the slack variables $x_2^{k,u}$ and $x_Q^{\tilde{k},u}$ we show the corresponding feasibility intervals.
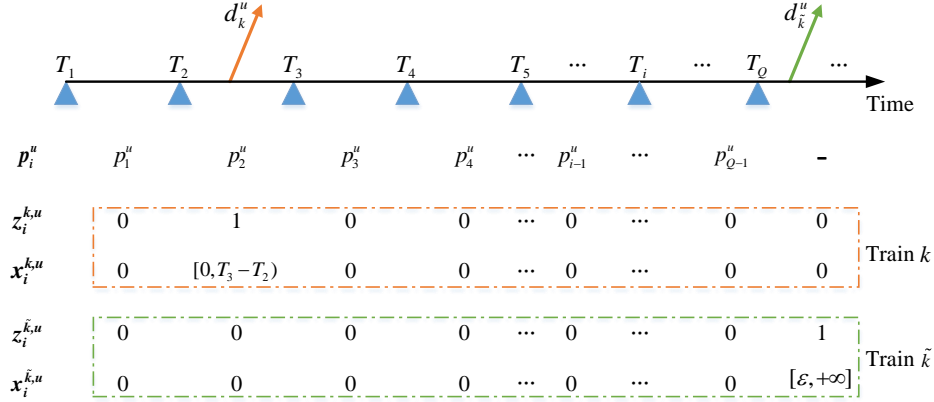
Figure 6: An example of feasible values for the coupling formulation

Modeling passenger arrivals. Observe first that in a period $i \in \mathcal{Q}$ without train departures from platform u, the number of passengers accumulating during the period amounts to $p_i^u(T_{i+1} - T_i)$.

We introduce an auxiliary binary variable $b_i^{k,u}$, which is 1 if train k departs from platform u in a period $j \le i$, that is:

$$b_i^{k,u} = \sum_{j=1}^{i} z_j^{k,u}, \quad u \in \mathcal{U}, k \in \mathcal{K}, i \in \mathcal{Q} \tag{20}$$

Clearly, if train $k-1$ leaves in a period before period i, and train k leaves in a period after i, then passengers arriving at the station accumulate on the platform during the entire period i, and the contribution of (the passengers arriving in) period i to the total queue amounts to $p_i^u(T_{i+1} - T_i)$, which is also equivalent to:

$$p_i^u(T_{i+1} - T_i)(b_i^{k-1,u} - b_i^{k,u}), \quad u \in \mathcal{U}, k \in \mathcal{K} \setminus \{1, K+1\}, i \in \mathcal{Q} \setminus \{Q\} \tag{21}$$

Now, if train $k-1$ departs in period j, then the contribution of period j to the total queue must be corrected according to the offset, and amounts to:

$$p_j^u(T_{j+1} - T_j)(b_j^{k-1,u} - b_j^{k,u}) - p_j^u x_j^{k-1,u} = p_j^u(T_{j+1} - T_j)(b_j^{k-1,u} - b_j^{k,u}) - \sum_{i=1}^{Q-1} p_i^u x_i^{k-1,u},$$
$$u \in \mathcal{U}, k \in \mathcal{K} \setminus \{1, K+1\}, j \in \mathcal{Q} \setminus \{Q\} \tag{22}$$

Finally, if train k leaves in a period l, the contribution of period l to the queue for train k will amount to

$$p_l^u x_l^{k,u} = \sum_{i=1}^{Q-1} p_i^u x_i^{k,u}, \quad u \in \mathcal{U}, k \in \mathcal{K}, l \in \mathcal{Q} \setminus \{Q\} \tag{23}$$

Combining all the contributions and remembering that the difference $(b_i^{k-1,u} - b_i^{k,u})$ is 1 precisely in the periods when $k-1$ has left u and k has not left yet, we have that the number of passengers waiting for train k at platform u is:

$$\sum_{i=1}^{Q-1} p_i^u(T_{i+1} - T_i)(b_i^{k-1,u} - b_i^{k,u}) - \sum_{i=1}^{Q-1} p_i^u x_i^{k-1,u} + \sum_{i=1}^{Q-1} p_i^u x_i^{k,u}, \quad u \in \mathcal{U}, k \in \mathcal{K} \setminus \{1, K+1\} \tag{24}$$

18

The quantity (24) replaces the term $p(d^u_{k-1}, d^u_k, u)$ when linearizing Equation (9).

3.5   Extending the models to multiple arrival rate samples

So far, we have introduced two MILP models, one for each strategy. One program modeling the ALTERNATE strategy refers to as ALTERNATEMILP, with constraints (1)∼(5), and another program modeling the SHORTTURN strategy refers to as SHORTTURNMILP, with constraints (1)∼(3) and (6)∼(7). The two models share the same objective function (15). Both models receive as input the $2N \times (Q-1)$ matrix P of the passenger arrival rates. Each model then computes the best timetable $\mathbf{d} \in \mathbb{R}^{\mathcal{K} \times \mathcal{U}}$, according to the strategy.

With a series of historical samples $\mathcal{P}^0 = \{P^1, \ldots, P^{RS}\}$ of the arrival rate matrix[10], we modify the MILP models for ALTERNATEMILP and SHORTTURNMILP for finding departures $\mathbf{d} \in \mathbb{R}^{\mathcal{K} \times \mathcal{U}}$ that are "good" with respect to all our samples. To this end, we proceed as follows. Note that, for the fixed timetable, by Remark 2, the value of the variables describing the passenger flows introduced in Section 3.2 depends only on the matrix P and are uniquely determined. In the extended model, for $r \in \mathcal{R}$, we introduce variables $pw^{u,r}_k$, $pb^{u,r}_k$, $pr^{u,r}_k$ and $pa^{u,r}_k$, to represent passengers waiting, boarding, onboard, and alighting train k at platform u when the arrival rate matrix is P^r, respectively. Every constraint from (8) to (12) is then substituted by RS copies, and in copy r, the original variables are replaced with the corresponding r-th extended counterpart, for $r \in \mathcal{R}$. Therefore, for instance, constraint (12) is replaced by the RS constraints

$$pr^{u,r}_k = pr^{u-1,r}_k + pb^{u,r}_k - pa^{u,r}_k, \quad k \in \mathcal{K}, u \in \mathcal{U} \setminus \{1, N+1\}, r \in \mathcal{R}$$

For the term $p(d^u_{k-1}, d^u_k, u)$, appearing in constraint (9), this will be replaced in the r-th copy of Equation (24) by the quantity

$$\sum_{i=1}^{Q-1} p^{u,r}_i (T_{i+1} - T_i)(b^{k-1,u}_i - b^{k,u}_i) - \sum_{i=1}^{Q-1} p^{u,r}_i x^{k-1,u}_i + \sum_{i=1}^{Q-1} p^{u,r}_i x^{k,u}_i, \quad u \in \mathcal{U}, k \in \mathcal{K} \setminus \{1, K+1\}, r \in \mathcal{R}$$

Note that Remark 2 also applies to the extended model. Therefore, let $\bar{\mathbf{d}}$ be a timetable, and let $\mathbf{pw}(\bar{\mathbf{d}}), \mathbf{pb}(\bar{\mathbf{d}}) \in \mathbb{R}^{|\mathcal{K}| \times |\mathcal{U}| \times RS}$ be the waiting passengers vector and the boarding passenger vector associated with $\bar{\mathbf{d}}$, respectively. Then, we denote by $f_R(\bar{\mathbf{d}})$ the objective function value associated with the timetable $\bar{\mathbf{d}}$, namely,

$$
\begin{aligned}
f_R(\bar{\mathbf{d}}) &= \lambda_w \cdot \frac{1}{RS} \sum_r f^r_w(\bar{\mathbf{d}}) + \lambda_h f_h(\bar{\mathbf{d}}) \\
&= \lambda_w \cdot \frac{1}{RS} \sum_r \sum_k \sum_u pw^{u,r}_k(\bar{\mathbf{d}}) - pb^{u,r}_k(\bar{\mathbf{d}}) + \lambda_h \sum_k \sum_u \max\{\bar{d}^u_k - \bar{d}^u_{k-1} - H_p, 0\} \quad (25)
\end{aligned}
$$

Remark 4. *Similarly for the other constraints and the full model, see Appendix B. Particularly, with a total of RS historical passenger samples $\mathcal{P}^0 = \{P^1, \ldots, P^{RS}\}$, the objective function will be applied in the two-stage approach (i.e., in Task 1) to find the appropriate recovery strategy and timetable for the variable passenger demands under the disruption.*

---

[10]There are a total of RS samples, denoted as $\mathcal{R} = \{1, 2, \ldots, RS\}$. Those samples could be obtained from different days of metro operation data to represent the variability of passenger demands.

4 Overall solution: a two-stage approach

With the historical and real-time detected data (like passenger demands and end time of the disruption), we develop a two-stage approach to provide adaptively optimized timetable so as to meet the real scenario as much as possible. The models linearized by the hybrid formulation (In Section 3) will be exploited in our approach to the recovery problem.

In Section 1, we have identified two phases between the start and the end of the blockage. In Phase I, the problem is detected and the recovery strategy is established and communicated to the system and operators. In Phase II, the recovery strategy is implemented. Accordingly, two-stage approach will split into two tasks: The first task is carried out in Phase I and is devoted to selecting the best strategy by comparing the timetables produced by the different models. The output of this first task will be the selected model, and the corresponding initial timetable that will apply to the next phase. Once the strategy is selected (and implemented by the dispatchers), we enter Phase II, and correspondingly, the second task of our approach starts. During the second task, we use the selected model to take real-time dispatching decisions. These decisions are taken by applying the selected model and by updating the entries of the sample matrices to their actual values for the past events. Clearly, in this task, the associated instances must be solved very quickly in order to provide updated plans in real-time scenarios.

First task. In this task, we need first to decide between the ALTERNATE and the SHORTTURN strategies. In the SHORTTURN strategy, we also need to decide where to place the short-cut crossover C_o (see Figure 4). Accordingly, we will solve one MILP for ALTERNATE and one MILP for SHORTTURN and each position of the crossover C_o. As input to this task, we have the set $\mathcal{P}^0 = \{P^1, \ldots, P^{RS}\}$ of historical arrival rate matrices, the prediction on the end time T_{end}^0 of the disruption, and all other relevant input data.

The strategy (and cross-over position in case of SHORTTURN) returning the best value of the extended objective function is selected at the end of this process. The associated model is denoted as BESTMILP, and the initially extended timetable is $\mathbf{d}^0$. Once the strategy is selected, the dispatcher will communicate it to the system and operators (see Figure 1) along with the initial timetable $\mathbf{d}^0$, which is input as incumbent solution $\mathbf{d}^* = \mathbf{d}^0$ to the next task. At this point, Phase II begins.

Second task. In this task, the model BESTMILP is solved iteratively in the real-time setting, for example, every ΔT seconds, starting at T_{sta}. At any new resolution, time moves forward, and some of the input parameters may change and some variables corresponding to past events are fixed to the actually realized values. Denote by $\tau^n = T_{sta} + n \cdot \Delta T$ the time iteration $n = 1, \ldots$ that is executed, and let BESTMILP^n be the model solved. In task 1, $\mathcal{P}^0 = \{P^1, \ldots, P^{RS}\}$ is the original set of sample matrices and T_{end}^0 is the predicted disruption end time. As the iterations proceed over time, however, better estimations may arrive and the predicted end time may change. We denote by T_{end}^n the end time predicted at iteration n. Similarly, at iteration n, all arrival rates before τ^n are known exactly. Therefore, we modify the matrices in $\mathcal{P}^0$ by substituting the original predicted values with the actually realized values. We denote by $\mathcal{P}^n$ the modified set of sample matrices at time τ^n. Additionally, all departure and arrival times before τ^n are fixed to their actual values and become constants in the current BESTMILP^n.

After solving BESTMILP^n at iteration n, the new solution (and associated timetable $\mathbf{d}^n$) is compared with the incumbent $\mathbf{d}^*$. In principle, the current timetable $\mathbf{d}^n$ is optimal and should replace the incumbent and be communicated to the dispatchers and the field through the signaling system. However, to enforce

some stability in the solution, the timetable is updated only if the objective value of this new solution $f(\mathbf{d}^n)$ is significantly better than the old one, namely, at least γ times the incumbent evaluated $f(\mathbf{d}^*)$ at time τ^n, where $\gamma \leq 1$ is a predefined parameter.
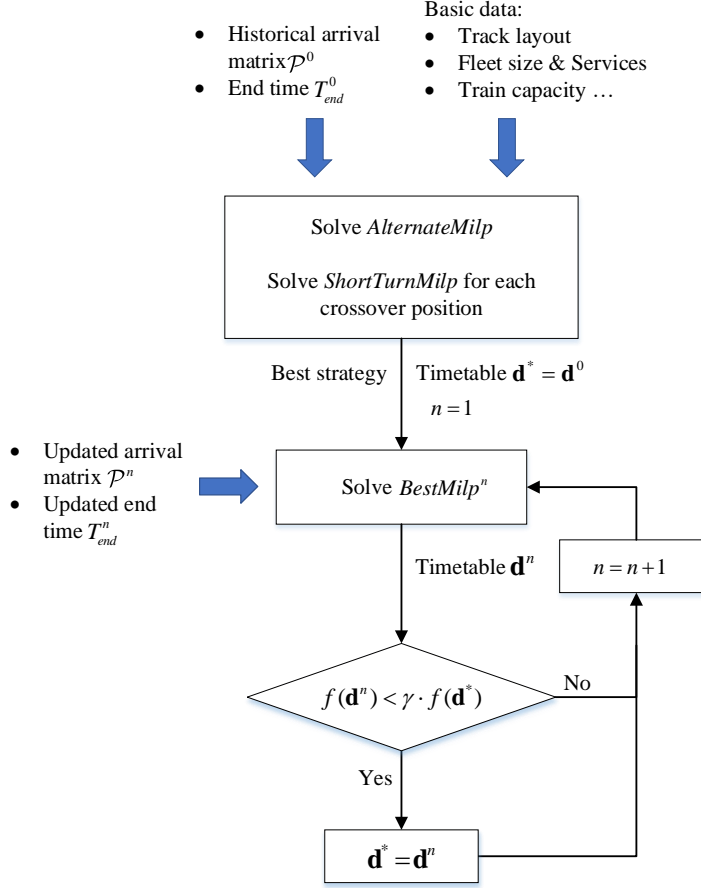

Figure 7: The flowchart of the two-stage approach

5 Numerical experiments

In this section, we present our numerical experiments. All instances are from Line 2 of the Beijing metro system. Our algorithm is coded in C# on an Intel(R) Core(TM) i5-7200 CPU @ 2.50 GHz with 16.00 GB RAM. As the MILP solver, the IBM CPLEX 12.7.1 is used to solve the proposed models with default settings.

5.1 The test-bed

Beijing metro Line 2 is a bidirectional loop line, consisting of 18 stations and 36 sections in total in the inner and outer loops (Figure 8). In the default operational mode, the inner loop is traversed clockwise starting at JISHUITAN, whereas the outer loop is traversed anticlockwise starting at XIZHIMEN. Beijing metro Line 2 has a rectangular shape around the city center. Additional data, including the platform index, the running time to the next station, the positions of the 7 crossover tracks, and the minimum dwell time are given in Appendix A. At a given point in time, passenger arrival rates are very different for the different stations. Similarly, for a given station, arrival rates vary significantly over the day. Real-life
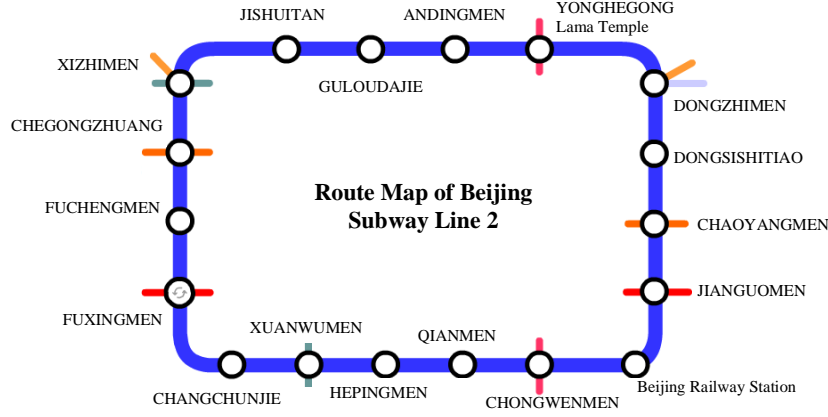
Figure 8: The route map of Beijing metro Line 2

passenger arrival rates[11] P^0 at different stations for a day are depicted in Figure 9. In the peak hour, the arrival rate even exceeds 60 persons/min at XIZHIMEN, JISHUITAN, DONGZHIMEN, and reaches 100 persons/min at Beijing Railway Station.
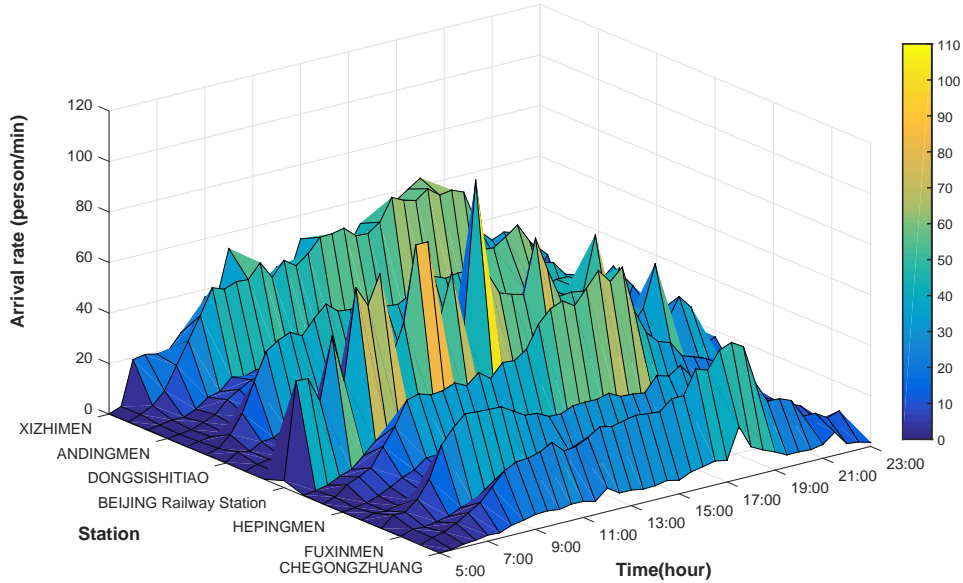

Figure 9: Passenger arrival rates at different stations of Beijing metro Line 2 in a weekend day

In our numerical experiments, we assume the disruption occurs at some time before 9:30 between CHANGCHUNJIE and FUXINGMEN in the inner direction. The disruption is initially estimated to end at 10:00. We then consider a 30-minute planning time horizon from 9:30 to 10:00 - assuming that Task 1 (and Phase I) can be completed no later than 9:30. The fleet amounts to 30 trains, providing 40 services during our planning horizon. The capacity C of the trains is 1200 passengers. Finally, we have a total of ten historical arrival rate samples, presented by $\mathcal{P}^0 = \{P^1, \ldots P^{10}\}$. When the disruption occurs, all in-service trains stop at the nearest stations and wait for dispatching directions. The second task (and Phase II) starts at 9:30.

---

[11]These are confidential, real-life data from Beijing Subway Company.

5.2   First task results

In this task, we solve seven instances of MILP models, one corresponding to the ALTERNATE strategy and six for the SHORTTURN strategy (namely, one for each candidate position of the short-turn crossover, denoted as $C_o = 1, \ldots, 6$, see Figure 4 and Table 7). For each instance, we fix the maximum CPU time to 150 seconds.

In Table 3, we show the results for task 1. The first line corresponds to the ALTERNATE strategy ("ALT" in the table), whereas the other lines correspond to SHORTTURN strategy ("TURN") for each crossover position (from $c1$ to $c6$). In the ALTERNATE strategy, the line is still divided into the inner and outer loop (I/O), whereas for the SHORTTURN strategy, we identify a left loop, a right loop and a bidirectional section (L/R/B). Especially when setting $c6$ (i.e., C_o overlaps the last crossover track C_b in Figure 4), the whole line is split into the rest and bidirectional section (Re/B). For all these segments, in Table 3, we give the number of platforms, the fleet size and the number of train services. Then, the average objective value $\bar{f}_R$, the average number of nonboarding passengers $\bar{f}_w$, the average headway deviation $\bar{f}_h$[12] are provided for each strategy with the given maximum allowed dwell time s_{max} (constraint (2)). Finally, in the last column, we display the integrality gap value returned by CPLEX at termination.

Table 3: Initial optimization results of two strategies

Strategy	# Platforms	Fleet size	# Services	s_{max}	$\bar{f}_R$	$\bar{f}_w$	$\bar{f}_h$	Gap(%)
ALT	18/18	15/15	17/16	690s	17.19	11.05	23.33	3.00%
TURN_c1	4/27/6	4/25/1	11/32/2	180s	4.79	5.79	3.79	13.86%
TURN_c2	6/25/6	6/23/1	13/29/2	180s	4.49	5.53	3.44	7.09%
TURN_c3	14/17/6	14/15/1	21/21/2	180s	4.89	6.01	3.77	12.20%
TURN_c4	20/11/6	19/10/1	26/17/2	180s	4.77	6.00	3.53	11.94%
TURN_c5	22/9/6	21/8/1	27/14/2	180s	4.85	6.26	3.44	15.63%
TURN_c6	31/6	29/1	36/2	180s	4.87	6.19	3.56	13.31%

The extended maximal dwell time (690 seconds) with the ALTERNATE strategy is a consequence of potential conflicts between trains running in opposite directions on the long shared track. This of course causes a substantial increase of headway deviation $\bar{f}_h$ and nonboarding passengers $\bar{f}_w$ with the cost of increased total travel time. In contrast, dwell time can be shortened to 180 seconds with SHORTTURN. SHORTTURN with crossover $c2$ ("TURN_c2"), with 6, 25 and 6 platforms in left, right and bidirectional segment and has the best objective value 4.49.

In Figure 10 we give some additional details of the solutions associated with "ALT", "TURN_c2" and "TURN_c6". In particular, we present the timetables (visualized as *train graphs*) and the total boarding and final nonboarding passengers[13]. In the figure, for ALTERNATE, the timetables for the inner and outer loop are represented by red and blue lines (the sequence of platforms is shown on the left and right y-axis, respectively). Longer stops are necessary to avoid conflicts on the shared outer platforms 22, 23 and 24, especially in the green rectangle in the dashed line. Observe the presence of nonboarding passengers in almost every station (mainly concentrated at the end of the disruption period). They sum up to more than 30% of all passengers traveling in the system, with a stunning 94% of the passengers stranded at platform 15. In contrast, the timetable in Figure 10(c), returned by the SHORTTURN strategy (in red, blue and orange the trains on the left, right and bidirectional segments, respectively) has a small headway deviation; the non-boarding passengers (13.8% of all passengers) are mainly concentrated at platforms 14, 15, 21 and 22, varying from 57% to 89% of the final nonboarding passengers at each platform. Merging

---

[12]Here, $\bar{f}_R$, $\bar{f}_w$ and $\bar{f}_h$ are the average values of the extended objective function f_R, nonboarding passengers f_w and headway deviation f_h (see Equations (25), (13) and (14)) for each train at each station on each sample, where the coefficients in the convex combination (15) are set as $\lambda_w = \lambda_h = 1/2$.

[13]Note that, those two passenger numbers are the average corresponding to a total of ten inputted passenger samples $\mathcal{P}^0$. Final boarding passengers mean the number of passengers that are not able to board before the end of the disruption.

5.3   Second task results

For this task, we allow 30 seconds of computing time for each iteration. Since passenger arrival rates are sampled every five minutes, in our experiments we also solve a new instance of BESTMILP every 5 minutes (i.e. $\Delta T = 300$ seconds). Additionally, since stability is not an issue in these tests, we let $\gamma = 1$.

According to the results presented in Table 3, the SHORTTURN strategy with crossover $c2$ (i.e., "TURN_c2") is the best and is therefore selected for the online rescheduling. Initially, the end of the planning horizon T_{end}^0 is set to 10:00 am, namely, the expected end time of the disruption. However, at 9:50, we simulate the decision of delaying T_{end}^4 to 10:05 due to a more accurate prediction on the disruption duration. Therefore, in total we have six iterations, i.e., $n = 1, \ldots, 6$.

The results of the second task iterations are presented in Table 4. To fully understand the figures in this table, keep in mind that at iteration n, i.e., at time $\tau^n = T_{sta} + n \cdot \Delta T$, we solve problem BESTMILP^n, with the current set $\mathcal{P}^n$ of partially realized sample matrices, current predicted end time T_{end}^n and a fixed timetable for arrivals and departures before τ^n. Every row (except the last) corresponds to an iteration of the algorithm, including the first task iteration 0. The objective value at iteration n is shown in the column $\bar{f}_R^n$, whereas $\bar{f}_w^n$ and $\bar{f}_h^n$ are the two components in the objective (before averaging), namely, the nonboarding passengers, and the headway component, respectively.

To understand the role of re-optimization, we also give in column $f^n(\mathbf{d}^{n-1})$ the value of the optimal timetable computed at iteration $n-1$, evaluated with the updated arrival rates $\mathcal{P}^n$ at iteration n. Note that at iteration 4, the new prediction on the end time becomes available and there is a sharp increase in the passenger component of the objective function value. This is because the expected duration of the disruption has increased, and consequently, also the number of nonboarding passengers. However, this does not correspond to more difficult MILPs for iteration 4, 5, 6, as the integrality gap at termination is constantly below 1%. Additionally, observe that from iteration 2, CPLEX can exploit the previous solution as a warm-start solution: this is clearly helping the computation, as the integral gap at termination falls from almost 5% to values of approximately 1%.

Additionally, we compute the first task initial timetable with the additional information on the extended duration of the disruption[16], namely, by letting $T_{end}^0 = 10:05$. Let this solution be $\mathbf{d}^{0*}$. We can now use the actually realized set $\mathcal{P}^6$ of arrival rates to evaluate. The resulting objective value is 9.08. Compared with the final $\bar{f}_R^6 = 7.85$, we see how the online recalculation allows for significant improvements over the initial solution.

Then, the final timetable is visualized in Figure 11(a), whereas Figure 11(b) shows how the objectives evolve over the iterations of the algorithm[17].

---

[14]For the bidirectional and rest-zone train, relevant routes respectively are $21 \to \cdots \to 24 \cdots \to 21$ and $16 \to \cdots \to 18 \to 1 \to \cdots \to 12 \to 24 \to \cdots \to 36 \to 19 \cdots \to 21$.

[15]Trains travelling through inner (outer) stop indices are represented with dotted (solid) lines, shown in left (right) y-axis.

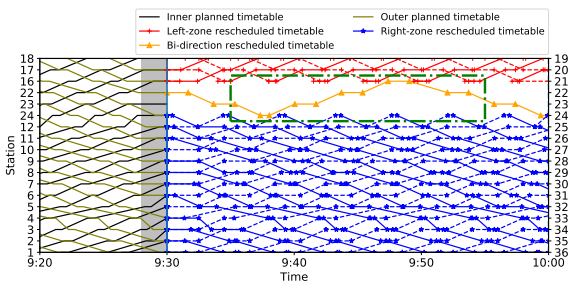[16]Observe that the initial timetable $\mathbf{d}^0$ is not feasible anymore since T_{end} is changed.

[17]In the figure, "total" represents the value of the whole transit line.
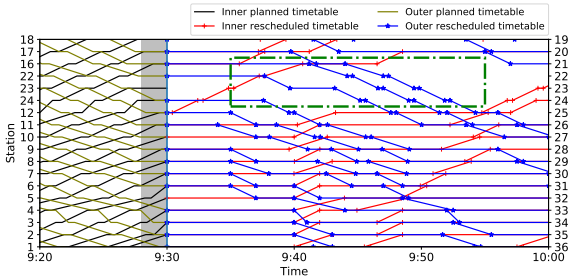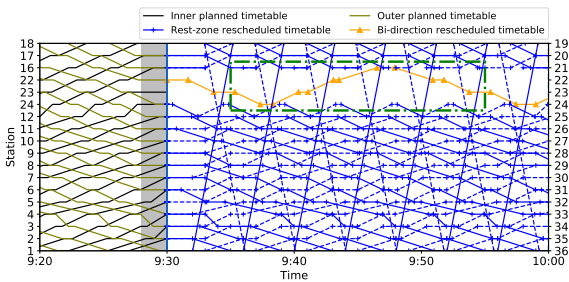
(a) Initial timetable with ALTERNATE

(b) Total boarding and final nonboarding passengers with ALTERNATE

(c) Initial timetable with SHORTTURN and crossover $c2^{15}$.

(d) Total boarding and final nonboarding passengers with SHORTTURN and crossover $c2$

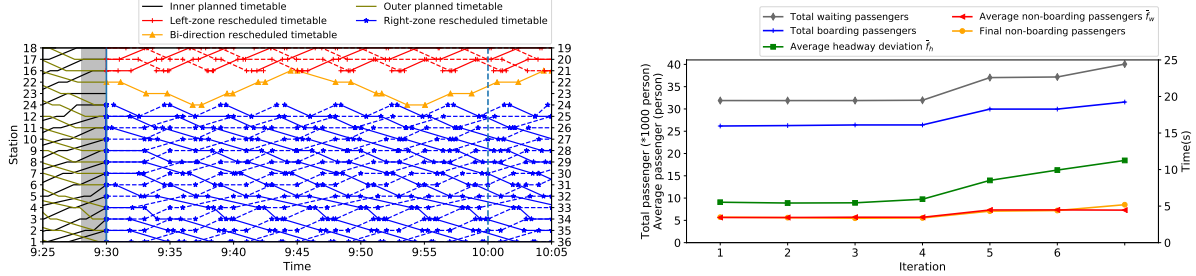(e) Initial timetable with SHORTTURN and crossover $c6^{15}$.

(f) Total boarding and final nonboarding passengers with SHORTTURN and crossover $c6$

Figure 10: Initial timetables and total boarding vs final nonboarding passengers associated with different strategies

Table 4: Computation results of strategy "Turn_c2" in the optimization process

Iter.	τ^n	T_{end}^n	$f^n(\mathbf{d}^{n-1})$	$\bar{f}_R^n$	$\bar{f}_w^n$	$\bar{f}_h^n$	Gap(%)
0	9:30	10:00	-	4.49	5.53	3.44	7.09%
1	9:35	10:00	4.48	4.43	5.42	3.44	4.74%
2	9:40	10:00	4.58	4.48	5.45	3.48	1.26%
3	9:45	10:00	4.72	4.70	5.94	3.47	0.21%
4	9:50	10:05	-	6.49	8.52	4.47	0.14%
5	9:55	10:05	7.22	7.20	9.93	4.48	0.91%
6	10:00	10:05	8.06	7.85	11.25	4.45	0.07%

(a) The online timetable obtained by ShortTurn and crossover $c2$

(b) Various costs behavior

Figure 11: The new timetable and relevant performance at the second task

5.4 Experiments with different weight factors

The relative weights of the passengers-related objective $\bar{f}_w$ and the headway-related objective $\bar{f}_h$ in the overall objective $\bar{f}_R$ are given by the coefficients λ_w, λ_h in the convex combination (15). Here, we study how $\bar{f}_R$ changes with such coefficients, in the case of the Alternate strategy (i.e., model AlternateMilp) and the ShortTurn strategy (i.e., model ShortTurnMilp) with crossover $c2$ at the first task.

For $\lambda_w \in [0,1]$, Figure 12(a) shows the behavior of average nonboarding passengers $\bar{f}_w$ and average headway deviation $\bar{f}_h$ for both strategies. With the increased weight factor of nonboarding passenger λ_w, the average nonboarding passengers $\bar{f}_w$ (solid blue and orange lines, value range on the left y-axis) drops significantly from 15 to 7.5 in the Alternate, and from 7.4 to 6.2 in the ShortTurn. In contrast, the average headway deviation $\bar{f}_h$ of the two strategies (represented by the dashed blue and orange lines, range on the right y-axis), increase from 21.1 to 31.7, and from 4.1 to 8.9, with 50.5% and 119% growth, respectively. Note that the impact of increasing λ_w is much stronger on the $\bar{f}_w$ component for the Alternate, mainly because by starting from a higher value, there is more room for improvements (by varying dwell times). Obviously, these two objectives $\bar{f}_w$ and $\bar{f}_h$ are not co-monotone with different tendencies of decline and growth, corresponding to a large number of waiting passengers on Line 2. With increasing λ_w, the cumulative effects of these two terms on the objective value $\bar{f}_R$ is shown in Figure 12(b).

5.5 Experiments with different arrival rates

Next, we move on to studying the influence of arrival rates on the solution and algorithm performance. To this end, starting from our family of arrival rate matrices $\mathcal{P}^0 = \{P^1, \ldots, P^{10}\}$, we construct 10 families $\mathcal{P}^0(j) = \{r_j \cdot P^1, \ldots, r_j \cdot P^{10}\}$, with $r_j = 0.2j$, $j = 1, \ldots, 10$. Then, we apply extended models AlternateMilp and ShortTurnMilp (with crossover c_2). The results for average nonboarding passengers $\bar{f}_w$, average headway deviation $\bar{f}_h$ and objective value $\bar{f}_R$ are shown in Figure 13, respectively. For the
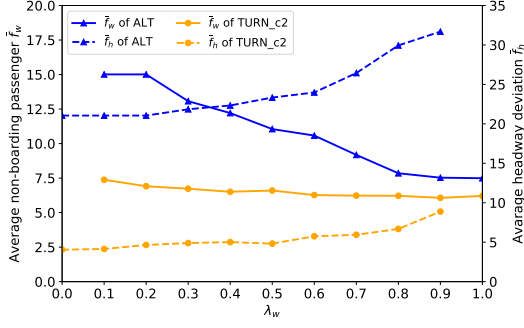
(a) Behavior of $\bar{f}_w$ and $\bar{f}_h$ with λ_w
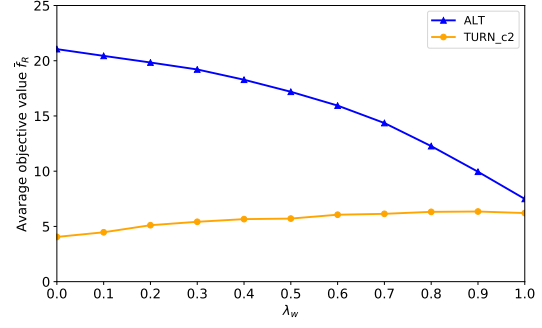
(b) Behavior of $\bar{f}_R$ with λ_w

Figure 12: The performance of ALTERNATES and SHORTTURN strategies with λ_w

low value of the coefficient, the two strategies have similar effects on $\bar{f}_w$. Then, the increasing number of passengers causes a larger growth of average nonboarding passengers $\bar{f}_w$ for the ALTERNATE compared with the SHORTTURN (Figure 13(a)). Indeed, due to the shared section, the ALTERNATE requires longer dwell time to avoid potential conflicts, leading to large headway deviations and, ultimately, to insufficient capacity to transport passengers. Trying to transport more passengers, both strategies will tend to exploit longer dwell times and this in turn induces a gradual increase in the average headway deviation $\bar{f}_h$. However, the increase of $\bar{f}_R$ is mainly determined by the corresponding increase of $\bar{f}_w$, see Figure 13(b).

Our experiments indicate that, in general, the SHORTTURN strategy (with the best possible choice for the border cross-over) dominates the ALTERNATE, at the cost of an increase in computational time. This increased time is, however, still within the acceptable time limits [18].
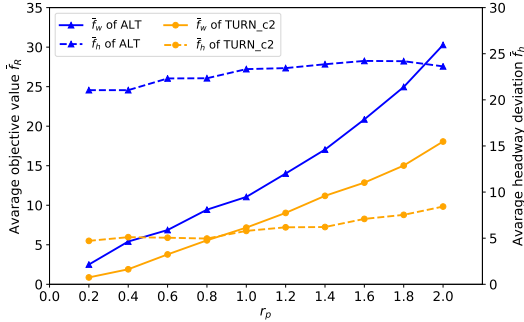
(a) Variation tendencies of $\bar{f}_w$ and $\bar{f}_h$ with r_p
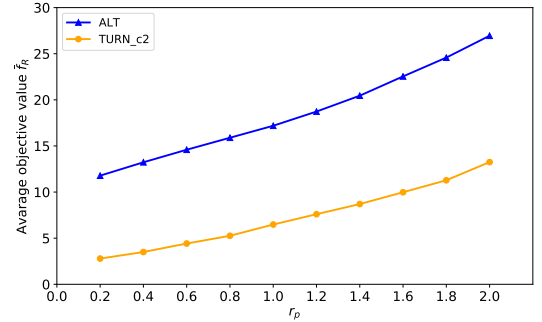
(b) Variation tendency of $\bar{f}_h$ with r_p

Figure 13: The performance of ALTERNATES and SHORTTURN strategies with λ_w

5.6 Experiments with "full" time-indexed formulation

For comparison purposes, we also develop and test a "full" time-indexed formulation for the ALTERNATE strategy. By "full" we mean that both passenger flows and train schedules are modeled through binary time-indexed variables. The results are given in Table 5. With larger time steps (15 secs and 30 secs), instances become infeasible. With a time step of 5 secs, CPLEX needs 6000 secs to reach a 24.89% gap,

---

[18]In practice, the average number of transfer process will be increased for passengers in the SHORTTURN strategy, but with less waiting time. Additionally, this strategy is a more flexible operation for dynamic passenger demands than the ALTERNATE strategy, while also requires the much higher automatic level of crossover switch.

due to the very large number of binary variables and constraints. Even if our implementation of the time-indexed formulation is rather straightforward, the long-running times reported in Table 5 indicate that it is not an immediate concern to develop a time-indexed formulation that is competitive with our mixed MILP model.

Table 5: Optimization result of full time-indexed formulation in the ALTERNATE strategy

Time step (seconds)	computation time (seconds)	$\bar{f}_R$	$\bar{f}_w$	$\bar{f}_h$	Gap
5	6000	17.67	12.17	23.16	24.89%
15	infeasible -	-	-	-	
30	infeasible -	-	-	-	

6 Conclusions and future research

When a blockage or disruption occurs in a metro line, the overall service is disturbed, and passengers may start queuing up in stations, especially at peak hours. In this paper, we propose a two-stage methodology to select a suitable recovery strategy (first task), and subsequently to implement it in an efficient way (second task). In both tasks, we solve one or more MILP problems that also take into account passenger flows by minimizing a measure of the total nonboarding passengers. To this end, we introduce a mixed-type formulation by intertwining a big-M block (mainly for the train scheduling part) with a time-indexed block (for modeling passenger flows). This idea allowed us to solve realistic instances of our problem from a major metro line in Beijing. In our experiments, we show that the approach is capable of selecting the best strategy (among those modeled) and then support the dispatchers in rescheduling trains during the disruption and the recovery phase.

There are some open questions and possible future developments. First, when the OD data are available, one may extend these models to cope with the actual numbers of boarding and alighting passengers at each station. Second, this paper only focuses on two possible recovery strategies; therefore, modeling other alternatives is a natural extension. Third, a blockage in the ring line also likely affects the connected radial lines in a way that depends on passenger detour decisions. It would be interesting to model such behavior and then predict and possibly control the impact on the other lines.

Acknowledgment

Appendix A Basic operational data of Line 2

In Table 6, we give some relevant figures for each station and each direction, namely, the platform index, and running time to the next station on the loop. The last column gives the planned dwell time at each station.

Table 6: Running time and dwell time of Beijing metro Line 2

| Station name | Inner loop | | Outer loop | | Dwell time |
	Platform	Running time	Platform	Running time	
JISHUITAN	1	180	36	190	35
GULOUDAJIE	2	120	35	180	40
ANDINGMEN	3	120	34	180	30
YONGHEGONG Lama Temple	4	240	33	60	30
DONGZHIMEN	5	120	32	240	30
DONGSISHITIAO	6	120	31	120	30
CHAOYANGMEN	7	180	30	120	40
JIANGUOMEN	8	120	29	180	40
Beijing Railway Station	9	180	28	120	40
CHONGWENMEN	10	120	27	180	35
QIANMEN	11	180	26	180	35
HEPINGMEN	12	60	25	120	35
XUANWUMEN	13	120	24	120	35
CHANGCHUNJIE	14	180	23	120	40
FUXINGMEN	15	180	22	120	40
FUCHENGMEN	16	120	21	180	35
CHEGONGZHUANG	17	130	20	120	40
XIZHIMEN	18	190	19	120	30

The locations of 7 crossover tracks are presented in Table 7, including the station name and the platform indexes on the inner and out loop lines. A crossover connects a section between two stations in the inner loop to a section between the same pair of stations in the outer loop. Each section is identified by the pair of adjacent platforms. Therefore, for instance, crossover track 3 (between YONGHEGONG Lama Temple and DONGZHIMEN) connects a point on section $(4,5)$ in the inner loop to a point on section $(32,33)$ in the outer loop.

Table 7: The stations nearby crossover tracks on Beijing metro Line 2

Track	Stations name	Inner platforms	Outer platforms
c1	(CHEGONGZHUANG, XIZHIMEN)	(17,18)	(19,20)
c2	(XIZHIMEN,JISHUITAN)	(18,1)	(19,36)
c3	(YONGHEGONG Lama Temple,DONGZHIMEN)	(4,5)	(32,33)
c4	(CHAOYANGMEN,JIANGUOMEN)	(7,8)	(29,30)
c5	(JIANGUOMEN,Beijing Railway Station)	(8,9)	(28,29)
c6	(HEPINGMEN,XUANWUMEN)	(12,13)	(24,25)
c7	(FUXINGMEN,FUCHENGMEN)	(15,16)	(21,22)

Appendix B   Complete formulation for the ALTERNATE strategy

According to Figure 3(a), we give an example to illustrate the whole formulation model of the ALTERNATE strategy. With the extended objective function (25) and constraints (see Section 3), the MILP model can be written as:

$$\min \ f_R(\bar{\mathbf{d}})$$

$$s.t. \ s_{min} \leq d_k^u - a_k^u \leq s_{max}, \quad u \in \mathcal{U}, k \in \mathcal{K} \setminus \{1, K+1\} \tag{26a}$$

$$a_k^{v(u)} - d_k^u = R^{u,v(u)}, \quad u \in \mathcal{U}, k \in \mathcal{K} \tag{26b}$$

$$d_k^u - d_{k-1}^u \geq H_{min}, \quad u \in \mathcal{U}, k \in \mathcal{K} \setminus \{1, K+1\} \tag{26c}$$

$$d_k^{2N-q} - a_{k-1}^{2N-q+1} \geq 0, \quad k \in \mathcal{K}_O \setminus \{K+1\} \tag{26d}$$

$$d_k^{2N-q+1} - a_{k-1}^{2N-q} \geq 0, \quad k \in \mathcal{K}_I \setminus \{1\} \tag{26e}$$

$$d_k^{q-1} - a_l^{2N-q+2} \geq \Theta_c - M y^{kl}, \quad k \in \mathcal{K}_{\mathcal{I}}, l \in \mathcal{K}_O \tag{26f}$$

$$d_l^{2N-q-1} - a_k^{q+2} \geq \Theta_c + M(y^{kl} - 1), \quad k \in \mathcal{K}_{\mathcal{I}}, l \in \mathcal{K}_O \tag{26g}$$

$$pb_k^{u,r} \leq R_b \cdot s_k^u, \quad u \in \mathcal{U}, k \in \mathcal{K}, r \in \mathcal{R} \tag{26h}$$

$$pb_k^{u,r} \leq pw_k^u, \quad u \in \mathcal{U}, k \in \mathcal{K}, r \in \mathcal{R} \tag{26i}$$

$$pb_k^{u,r} \leq pc_k^u, \quad u \in \mathcal{U}, k \in \mathcal{K}, r \in \mathcal{R} \tag{26j}$$

$$\sum_{i=1}^{Q} z_i^{k,u} = 1, \quad u \in \mathcal{U}, k \in \mathcal{K} \tag{26k}$$

$$pw_k^{u,r} \geq [pw_{k-1}^{u,r} - pb_{k-1}^{u,r} + p(d_{k-1}^u, d_k^u, u)] - M z_Q^{k,u}, \quad u \in \mathcal{U}, k \in \mathcal{K} \setminus \{1, K+1\}, r \in \mathcal{R} \tag{26l}$$

$$p(d_{k-1}^u, d_k^u, u) = \sum_{i=1}^{Q-1} p_i^{u,r}(T_{i+1} - T_i)(b_i^{k-1,u} - b_i^{k,u}) - \sum_{i=1}^{Q-1} p_i^{u,r} x_i^{k-1,u} + \sum_{i=1}^{Q-1} p_i^{u,r} x_i^{k,u}, \tag{26m}$$
$$u \in \mathcal{U}, k \in \mathcal{K} \setminus \{1, K+1\}, r \in \mathcal{R}$$

$$b_i^{k,u} = \sum_{j=1}^{i} z_j^{k,u}, \quad u \in \mathcal{U}, k \in \mathcal{K}, i \in \mathcal{Q} \tag{26n}$$

$$d_k^u = \sum_{i=1}^{Q} T_i \cdot z_i^{k,u} + x_i^{k,u}, \quad u \in \mathcal{U}, k \in \mathcal{K} \tag{26o}$$

$$x_i^{k,u} \leq (T_{i+1} - T_i - \varepsilon) z_i^{k,u}, \quad u \in \mathcal{U}, k \in \mathcal{K}, i \in \mathcal{Q} \tag{26p}$$

$$x_i^{k,u} \geq \varepsilon \tag{26q}$$

$$z_i^{k,u} \in \{0,1\}, \quad u \in \mathcal{U}, k \in \mathcal{K}, i \in \mathcal{Q} \tag{26r}$$

$$d_k^u, a_k^u, pb_k^{u,r}, pw_k^{u,r} \in \mathbb{R}_+, \quad u \in \mathcal{U}, k \in \mathcal{K} \tag{26s}$$

Note that, constraints (26a)$\sim$(26g) guarantee train safety operations, including timetable constraints, disruption dependent constraints and sequencing constraints. The next seven constraints (26h)$\sim$(26n) are to linearize the $\bar{f}_w^r$ of the objection function. Finally, the last five constraints denote the relationship and values of those relevant variables.

References

[1] Eva Barrena, David Canca, Leandro C. Coelho, and Gilbert Laporte. Single-line rail rapid transit timetabling under dynamic passenger demand. *Transportation Research Part B: Methodological*, 70(Supplement C):134–150, 2014.

[2] Beijing Subway. `http://www.weibo.com/bjsubway` (in Chinese), July and August 2017.

[3] Stefan Binder, Yousef Maknoon, and Michel Bierlaire. The multi-objective railway timetable rescheduling problem. *Transportation Research Part C: Emerging Technologies*, 78:78–94, 2017.

[4] Valentina Cacchiani, Alberto Caprara, and Paolo Toth. Scheduling extra freight trains on railway networks. *Transportation Research Part B Methodological*, 44(2):215–231, 2010.

[5] Alberto Caprara, Matteo Fischetti, and Paolo Toth. Modeling and solving the train timetabling problem. *Operations research*, 50(5):851–861, 2002.

[6] Alberto Caprara, Michele Monaci, Paolo Toth, and Pier Luigi Guida. A lagrangian heuristic algorithm for a real-world train timetabling problem. *Discrete applied mathematics*, 154(5):738–753, 2006.

[7] Francesco Corman, Andrea D'Ariano, and Ingo A Hansen. Evaluating disturbance robustness of railway schedules. *Journal of Intelligent Transportation Systems*, 18(1):106–120, 2014.

[8] Francesco Corman, Andrea D'Ariano, Alessio D Marra, Dario Pacciarelli, and Marcella Samá. Integrating train scheduling and delay management in real-time railway traffic control. *Transportation Research Part E: Logistics and Transportation Review*, 105:213–239, 2017.

[9] J Cury, F Gomide, and M Mendes. A methodology for generation of optimal schedules for an underground railway system. *IEEE Transactions on automatic control*, 25(2):217–222, 1980.

[10] Andrea D'ariano, Dario Pacciarelli, and Marco Pranzo. A branch and bound algorithm for scheduling trains in a railway network. *European Journal of Operational Research*, 183(2):643–657, 2007.

[11] Xu Jun Eberlein, Nigel HM Wilson, and David Bernstein. Modeling real-time control strategies in public transit operations. In *Computer-aided transit scheduling*, pages 325–346. Springer, 1999.

[12] Yuan Gao, Leo Kroon, Marie Schmidt, and Lixing Yang. Rescheduling a metro line in an overcrowded situation after disruptions. *Transportation Research Part B Methodological*, 93:425–449, 2016.

[13] Nadjla Ghaemi, Oded Cats, and Rob MP Goverde. A microscopic model for optimal train shortturnings during complete blockages. *Transportation Research Part B: Methodological*, 105:423–437, 2017.

[14] Nadjla Ghaemi, Oded Cats, and Rob MP Goverde. Macroscopic multiple-station short-turning model in case of complete railway blockages. *Transportation Research Part C: Emerging Technologies*, 89:113–132, 2018.

[15] Ahmet B Keha, Ketan Khowala, and John W Fowler. Mixed integer programming formulations for single machine scheduling problems. *Computers & Industrial Engineering*, 56(1):357–367, 2009.

[16] Leonardo Lamorgese and Carlo Mannino. An exact decomposition approach for the real-time train dispatching problem. *Operations Research*, 63(1):48–64, 2015.

[17] Leonardo Lamorgese and Carlo Mannino. A non-compact formulation for job-shop scheduling problems in transportation. *Operations Research*, 2018. to appear.

[18] Rune Larsen, Marco Pranzo, Andrea DâĂŹAriano, Francesco Corman, and Dario Pacciarelli. Susceptibility of optimal train schedules to stochastic disturbances of process times. *Flexible Services and Manufacturing Journal*, 26(4):466–489, 2014.

[19] Shukai Li, Lixing Yang, and Ziyou Gao. Efficient real-time control design for automatic train regulation of metro loop lines. *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[20] Ilse Louwerse and Dennis Huisman. Adjusting a railway timetable in case of partial or complete blockades. *European Journal of Operational Research*, 235(3):583–593, 2014.

[21] Carlo Mannino and Alessandro Mascis. Optimal real-time traffic control in metro stations. *Operations Research*, 57(4):1026–1039, 2009.

[22] Carlo Mannino and Giorgio Sartor. The path & cycle formulation for the hotspot problem in air traffic management. In *18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[23] Alessandro Mascis and Dario Pacciarelli. Job-shop scheduling with blocking and no-wait constraints. *European Journal of Operational Research*, 143(3):498–517, 2002.

[24] Huimin Niu, Xuesong Zhou, and Ruhu Gao. Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints. *Transportation Research Part B Methodological*, 76:117–135, 2015.

[25] Susan Wynne O'Dell. *Optimal control strategies for a rail transit line*. PhD thesis, Massachusetts Institute of Technology, 1997.

[26] Brendan Pender, Graham Currie, Alexa Delbosc, and Nirajan Shiwakoti. Disruption recovery in passenger railways. *Transportation Research Record: Journal of the Transportation Research Board*, 2353:22–32, 2013.

[27] People's Daily. `http://finance.people.com.cn/n1/2018/0611/c1004-30048964.html` (in Chinese), June 2018.

[28] Michael L Pinedo. *Scheduling: theory, algorithms, and systems*. Springer, 2016.

[29] André Puong and Nigel HM Wilson. A train holding model for urban rail transit systems. In *Computer-aided systems in public transport*, pages 319–337. Springer, 2008.

[30] Maurice Queyranne and Andreas S Schulz. *Polyhedral approaches to machine scheduling*. TU, Fachbereich 3, Berlin, 1994.

[31] Marcella Samà, Francesco Corman, Dario Pacciarelli, et al. A variable neighbourhood search for fast train scheduling and routing during disturbed railway traffic situations. *Computers & Operations Research*, 78:480–499, 2017.

[32] Marcella Sama, Paola Pellegrini, Andrea D'Ariano, Joaquin Rodriguez, and Dario Pacciarelli. Ant colony optimization for the real-time train routing selection problem. *Transportation Research Part B: Methodological*, 85:89–108, 2016.

[33] Gabriel E Sánchez-Martínez, Haris N Koutsopoulos, and Nigel HM Wilson. Event-driven holding control for high-frequency transit. *Transportation Research Record: Journal of the Transportation Research Board*, (2535):65–72, 2015.

[34] Jan Dirk Schmocker, Shoshana Cooper, and William Adeney. Metro service delay recovery: Comparison of strategies and constraints across systems. *Transportation Research Record Journal of the Transportation Research Board*, 1930(1):30–37, 2005.

[35] Yihui Wang, Andrea D'Ariano, Jiateng Yin, Lingyun Meng, Tao Tang, and Bin Ning. Passenger demand oriented train scheduling and rolling stock circulation planning for an urban rail transit line. *Transportation Research Part B: Methodological*, 118:193–227, 2018.

[36] Xinhua News. `http://www.xinhuanet.com/politics/2015-09/26/c_128269673.htm` (in Chinese), September 2015.

[37] Xiaoming Xu, Keping Li, and Lixing Yang. Rescheduling subway trains by a discrete event model considering service balance performance. *Applied Mathematical Modelling*, 40(2):1446–1466, 2016.

[38] Jiateng Yin, Tao Tang, Lixing Yang, Ziyou Gao, and Bin Ran. Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: An approximate dynamic programming approach. *Transportation Research Part B Methodological*, 91:178–210, 2016.

[39] Jiateng Yin, Lixing Yang, Tao Tang, Ziyou Gao, and Bin Ran. Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches. *Transportation Research Part B: Methodological*, 97:182–213, 2017.

[40] Huimin Zhang, Shukai Li, and Lixing Yang. Real-time optimal train regulation design for metro lines with energy-saving. *Computers & Industrial Engineering*, 2018.

CRediT authorship contribution statement

Yeran Huang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Visualization, Writing - Original Draft, Writing - review & editing. **Carlo Mannino**: Methodology, Data Curation, Investigation, Formal analysis, Writing - review & editing, Project administration. **Lixing Yang**: Conceptualization, Investigation, Writing - review & editing, Project administration, Funding acquisition. **Tao Tang**: Resources, Project administration, Funding acquisition, Supervision, Writing - review & editing