Robust Path Recommendations During Public Transit Disruptions Under Demand Uncertainty

Baichuan Mo[a,*], Haris N. Koutsopoulos[b], Zuo-Jun Max Shen[c], Jinhua Zhao[d]

[a]*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139*
[b]*Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115*
[c]*Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, CA 94720*
[d]*Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 20139*

Abstract

When there are significant service disruptions in public transit systems, passengers usually need guidance to find alternative paths. This paper proposes a path recommendation model to mitigate congestion during public transit disruptions. Passengers with different origins, destinations, and departure times are recommended with different paths such that the system travel time is minimized. We model the path recommendation problem as an optimal flow problem with uncertain demand information. To tackle the lack of analytical formulation of travel times due to capacity constraints, we propose a simulation-based first-order approximation to transform the original problem into a linear program. Uncertainties in demand are modeled using robust optimization to protect the path recommendation strategies against inaccurate estimates. A real-world rail disruption scenario in the Chicago Transit Authority (CTA) system is used as a case study. Results show that even without considering uncertainty, the nominal model can reduce the system travel time by 9.1% (compared to the status quo), and outperforms the benchmark capacity-based path recommendation. The average travel time of passengers in the incident line (i.e., passengers receiving recommendations) is reduced more (-20.6% compared to the status quo). After incorporating the demand uncertainty, the robust model can further reduce system travel times. The best robust model can decrease the average travel time of incident-line passengers by 2.91% compared to the nominal model. The improvement of robust models is more prominent when the actual demand pattern is close to the worst-case demand.

Keywords: Path recommendation; Robust optimization; Rail disruptions; Demand uncertainty

1. Introduction

1.1. Background

Public transit (PT) systems play an important role in urban mobility. However, with aging systems, continuous expansion, and near-capacity operations, service disruptions often occur. These incidents may result in delays, cancellation of trips, and economic losses (Cox et al., 2011).

This study considers significant service disruptions in public transit systems where the service (or line/route) is interrupted for a relatively long period of time (e.g., 1 hour). During a disruption, affected passengers need to find an alternative path or use other travel modes (such as transfer to another bus route). However, due to a lack of knowledge of the system state (especially during incident time), the alternative

*Corresponding author

routes chosen by passengers may not be optimal or even cause more congestion (Mo et al., 2022b). For example, during a rail disruption, most of the passengers may choose bus routes that are parallel to the interrupted rail line as an alternative. However, given the limited capacity of buses, the parallel bus line may be over-saturated and passengers have to wait for a long time to board due to being denied boarding (or left behind).

1.2. Objectives and Challenges

One of the strategies to better guide passengers is to provide path recommendations so that the passenger flows are re-distributed in a better way and the system travel times are reduced. This can be seen as solving an **optimal passenger flow distribution (or assignment) problem** over a public transit network. However, there are several challenges to this problem.

- First, as the objective is to reduce the system travel time, an analytical formulation to calculate passengers' travel times is needed. However, a passenger's waiting times at the boarding and transfer stations are not only determined by other waiting passengers but also those who already boarded the same line as they reduce the vehicle's capacity (De Cea and Fernández, 1993). This complicated interaction makes it difficult to have an analytical formulation for passengers' travel time when the left behind is not negligible (which is usually the case during service disruptions). More details on this challenge are elaborated in Section 2.4.

- Second, there are many uncertainties in the system, such as the number of passengers using the PT system during incidents (i.e., demand uncertainty), incident duration, and whether passengers would follow the recommendations or not (i.e., behavior uncertainty). Previous studies have not considered uncertainties in modeling an optimal passenger flow problem.

This study aims to propose a path recommendation model to reduce crowding during public transit disruptions, also taking into account uncertainties due to inaccurate demand estimates. Different from previous recommendation systems that focus on maximizing individual preferences, this study targets a system objective by minimizing the total travel time of all passengers (including those who are not in the incident line/area). To address the aforementioned first challenge, we propose a simulation-based linearization to convert the total system travel time to a linear function of path flows using a first-order approximation, which leads to a tractable optimization problem. For the second challenge, this study focuses on the demand uncertainty (i.e., how many passengers will use the PT system during a service disruption) and models it within the robust optimization (RO) framework. The proposed approach is applied in a case study using data from the Chicago Transit Authority (CTA) system during a real-world urban rail disruption.

The main contributions of this paper are as follows:

- To tackle the non-analytical system travel time calculation, we propose a simulation-based linearization to convert the total system travel time to a linear function of path flows using first-order approximation. Importantly, we utilize the physical interaction between passengers and vehicles in a public transit system to efficiently calculate the gradient (i.e., marginal change of travel time) without running the simulation multiple times (as opposed to traditional black-box optimization).

- We use RO to model the demand uncertainty which protects the model against inaccurate demand estimation. Specifically, we derive the closed-form robust counterpart with respect to the intersection of one ellipsoidal and three polyhedral uncertainty sets. These uncertainties capture the demand

variations and the potential demand reduction during an incident. We also provide a feasible way of combining historical and survey data to quantify the uncertainty parameters.

The remainder of this paper is organized as follows. The literature review is presented in Section 2. In Section 3, we describe the problem and discuss the solution methods. Section 4 discusses model extensions and generalizability. We apply the proposed framework to the CTA system as a case study in Section 5. The model results are analyzed in Section 6. Finally, we conclude the paper and summarize the main findings in Section 7.

2. Literature review

2.1. Supply-side incident management

During a disruption, transit operators usually need to adjust services such as re-schedule timetables, re-route services, or design shuttle buses. Jespersen-Groth et al. (2009) mention that the disruption management process often involves solving three interrelated problems sequentially: timetable adjustment, rolling stock rescheduling, and crew rescheduling. These are supply-side incident management strategies that are different from path recommendations (demand side). Supply-side strategies are widely explored in the literature. For example, timetable rescheduling has been explored from both train-oriented (D'Ariano et al., 2008; D'Ariano and Pranzo, 2009; Corman et al., 2010, 2012, 2014; Louwerse and Huisman, 2014; Zhan et al., 2015) and passenger-oriented (Schöbel, 2007; Schachtebeck and Schöbel, 2010; Dollevoet et al., 2012; Kroon et al., 2015; Gao et al., 2016) aspects, where the former pays more attention to the details of the rail system and the handling of disruptions or disturbances, focusing on minimizing the delays of trains or the number of canceled trains. The latter aims at minimizing passengers' total delay after a disruption or disturbance. For shuttle bus designs, Kepaptsoglou and Karlaftis (2009) propose a methodological framework for planning and designing an efficient bus bridging network. Jin et al. (2016) use a column generation procedure to dynamically generate demand-responsive candidate bus routes for shuttle bus design. A more comprehensive review of supply-side recovery models and algorithms for real-time railway disturbance and disruption management can be found in Cacchiani et al. (2014).

2.2. Path recommendations during incidents

Most previous studies on path recommendations under incidents were conducted at a single OD level. That is, the main objective is to find available routes or the shortest path given an OD pair when the network is interrupted by incidents. For example, Bruglieri et al. (2015) designed a trip planner to find the fastest path in the public transit network during service disruptions based on real-time mobility information. Böhmová et al. (2013) developed a routing algorithm in urban public transportation to find reliable journeys that are robust against system delays. Roelofsen et al. (2018) provided a framework for generating and assessing alternative routes in case of disruptions in urban public transport systems. To the best of the authors' knowledge, none of the previous studies have considered path recommendations at the system level, that is, providing path recommendations for passengers of different OD pairs and with different departure times so that the system travel time is reduced.

2.3. Passenger evacuation under emergencies

Providing path recommendations during disruptions is related to the topic of passenger evacuation under emergencies. The objective of evacuation is usually to minimize the total evacuation time. In general, these

papers can be categorized into micro-level and macro-level based on how passenger flows are modeled and the spatial scope of the study area.

The micro-level studies usually use an agent-based simulation model to evaluate different evacuation strategies within some infrastructure. For example, Wang et al. (2013) simulated passenger evacuation under a fire emergency in Metro stations. Chen et al. (2017) developed four modeling approaches including a queuing model and an agent-based simulation to calculate the evacuation time under different emergency situations and evacuation plans. Hassannayebi et al. (2020) used an agent-based and discrete-event simulation model to assess the service level performance and crowdedness in a metro station under various disruption scenarios (e.g., train failure in the tunnel and fire at the station gallery). Zhou et al. (2019) proposed a hybrid bi-level model to optimize the number and initial locations of leaders who guide passengers' evacuation in urban rail transit stations during an evacuation.

The macro-level studies consider a larger study area (e.g., city-level) and aim to evacuate passengers from the incident area through various transportation modes. For example, Abdelgawad and Abdulhai (2012) developed an evacuation model to determine the routing and scheduling of subway and bus transit services used to alleviate congestion pressure during the evacuation of busy urban areas. Wang et al. (2019a) proposed an optimal bus bridging design method under operational disruptions on a single metro line. Tan et al. (2020) proposes an evacuation model with urban bus networks as alternatives in the case of common metro service disruptions by jointly designing the bus lines and frequencies.

The macro-level passenger evacuation is similar to the setup of this study, but with the following major differences. First, in our paper, the service disruption is not as severe as an emergency situation. The service will recover after a period of time and passengers are allowed to wait at a station. They do not necessarily need to cancel trips or follow evacuation plans as required in evacuation studies. Second, in this study, we assume that the service adjustment is known. The focus is on providing information to passengers to better utilize the existing resources/capacities of the system (demand side). However, the evacuation studies, since usually assuming the whole system breaks down, mainly focus on designing new services, such as routing and re-scheduling (supply side).

2.4. Travel time calculation in public transit networks

Passengers' travel time has two components: in-vehicle time and waiting time. In-vehicle time is not affected by passenger flows once passengers are onboard, thus is easy to model (e.g., modeled as a constant). However, the waiting time is more complicated to calculate if the system is congested with left behind due to capacity constraints.

Passengers' travel time is usually modeled in the context of transit assignment, using two major approaches: frequency-based (static) and schedule-based (dynamic). In the frequency-based transit assignment approach, the waiting time is either assumed to be inversely proportional to the (effective) service frequency (Wu et al., 1994; Schmöcker et al., 2011; Nielsen, 2000), or modeled as a congestion function (e.g., BRP) of previously boarded flows and new arrival flows with exogenously-calibrated parameters (De Cea and Fernández, 1993). The former method does not consider the left behind, and the latter only outputs a generalized waiting cost (rather than the waiting time as the vehicle capacity is not explicitly modeled) and requires a dedicated calibration process. Therefore, the frequency-based transit assignment model is not suitable for this study because congestion and left behind are not negligible during disruptions.

In terms of the schedule-based models (Nguyen et al., 2001; Hamdouch and Lawphongpanich, 2008; Hamdouch et al., 2014; Schmöcker et al., 2008), the waiting time can only be obtained after a dynamic

network loading (or simulation) process. For example, Schmöcker et al. (2008) used the fail-to-board probability to model the left behind. This probability is updated after each network loading and can be used to calculate the waiting time. However, in this way, the waiting time is still constant within each iteration. There is no direct way to formulate waiting time as a function of path flows.

Since formulating travel time as a function of path flows remains a challenge, the optimal passenger flow distribution in transit networks has no closed-form formulation. This study proposes a simulation-based first-order approximation to solve the original problem iteratively. With the proposed tractable linear programming model, uncertainties can also be incorporated.

2.5. Passenger queuing in over-saturated scenarios

Since the difficulty of travel time calculation arises from the waiting time due to being left behind, we also review previous studies on modeling passenger queuing in over-saturated scenarios. Passenger left behind is usually modeled by the following nonlinear constraint:

$$\text{Num boarding passengers} = \min\{\text{Num waiting passengers}, \text{Remaining capacity}\} \tag{1}$$

This constraint is resolved by the following methods in the literature: 1) transferring to a linear constraint with binary decision variables then solved by heuristics or other algorithms (Gao et al., 2016; Shi and Li, 2021), 2) meta-heuristics (e.g., genetic algorithm (GA), sequential quadratic programming) (Yang et al., 2012; Wang et al., 2015b), 3) iterative convex programming (Wang et al., 2015a), 4) approximate dynamic programming (Yin et al., 2016; Shi and Li, 2021), 5) effective passenger loading time period (with binary decision variables) (Niu and Zhou, 2013). Among these methods, modeling with binary decision variables requires solving large-scale integer programming (the number of decision variables equal to the number of time intervals times the number of platforms). This is usually solved by some heuristics and may not be applicable in a large-scale network. Another category of meta-heuristics methods (like GA) is not efficient and does not well utilize domain-specific knowledge. Specifically, iterative convex programming is slightly similar to our method. In each iteration, the approach fixes the number of waiting passengers and onboard passengers based on the timetable from the last iteration and a simulation model. In our study, we also have a fixed "flow pattern" from the last iteration, but we also capture the "marginal change" in flows using a first-order approximation (see Section 3.3 for details).

2.6. Simulation-based optimization

Simulation-based optimization methods are designed to solve optimization problems where the objective function and its derivatives are difficult and expensive to evaluate. These methods have been widely used to solve the problems of congestion pricing (Chen et al., 2016; He et al., 2017), traffic signal control (Osorio and Bierlaire, 2013; Osorio and Nanduri, 2015b,a; Chong and Osorio, 2018), transit scheduling (Zhang et al., 2017), route choice estimation (Mo et al., 2021, 2022a), ride-sharing (Cardin et al., 2017), supply chain management (Noordhoek et al., 2018), liner shipping (Dong and Song, 2009) and more. In general, there are three classes of methods for the SBO, including the direct search method, the gradient-based method, and the response surface (meta-model) method (Osorio and Bierlaire, 2013). In this paper, the proposed simulation-based first-order approximation is similar to a combination of the gradient-based and response surface (meta-model) methods. Specifically, we use the first-order approximation as a meta-model for the original objective function. In terms of the gradient calculation, instead of calling the simulation multiple times for the gradient evaluation (e.g., simultaneous perturbation stochastic approximation (Spall, 1997)),

we propose an efficient way to calculate the gradient based on its physical meaning. This greatly saves computational time compared to typical gradient-based methods.

2.7. Robust optimization (RO)

RO is a common approach to handling data uncertainty in optimization problems. RO generally needs to first specify a scope of some uncertain parameters. The scope is referred to as the "uncertainty set". The optimization problem is conducted over the worst-case realizations within the specified uncertainty set. This method is suitable for applications where there are uncertainties related to the model input parameters and when uncertainties can lead to significant penalties or infeasibility in practice. Since the solutions are optimal under the worst-case scenario, we treat the outputs of RO as a robust solution.

The solution method for RO problems involves generating a deterministic equivalent formulation, called the robust counterpart. Computational tractability of the robust counterpart has been a major practical difficulty (Ben-Tal et al., 2009). A variety of uncertainty sets have been identified for which the robust counterpart is reasonably tractable (Bertsimas et al., 2011).

The studies on RO have grown substantially over the past decades. Seminal papers include (Ben-Tal and Nemirovski, 1998), (Ben-Tal and Nemirovski, 1999) and (Bertsimas and Sim, 2004). Comprehensive surveys on the early literature can be found in Ben-Tal et al. (2009) and Bertsimas et al. (2011). The development of the RO methodology has allowed researchers to tackle problems with data uncertainty in a range of fields. Examples include renewable energy network design (Xiong et al., 2016), supply chain operations (Ma et al., 2018), health care logistics (Wang et al., 2019b), and ride-hailing (Guo et al., 2021).

However, to the best of the authors' knowledge, no existing papers have incorporated RO techniques into path recommendations during service disruptions. This research gap is important to address given the potentially inaccurate estimates of demand in public transit networks during an incident.

3. Methodology

3.1. Event-based public transit simulator

Before introducing the path recommendation, we first describe an event-based public transit simulator that is used across this study (Mo et al., 2020), especially for simulation-based linearization.

3.1.1. Simulator design

Figure 1 summarizes the main structure of the simulator. The inputs for the simulator are time-dependent OD demand (or smart card data), path shares, network structure, and train movement data (or timetable). Three objects are defined: trains, queues, and passengers. Trains are characterized by routes, train ID, current locations, and capacities. Passengers are queued based on their arrival times. Three different types of passengers are represented: left-behind passengers who were denied boarding from previous trains, new tap-in passengers from outside the system, and new transfer passengers from other lines. The left-behind passengers are usually at the head of the queue.

An event-based modeling framework is used to load the passengers onto the network. Two types of events are considered: train arrivals and train departures. The events are sorted by time and processed sequentially until all events are successfully completed during the analysis period. Train event lists (arrivals and departures) are generated according to the actual train movement data or timetable. Each event contains a train ID, occurrence time, and location (platform). Passengers are assigned to a path based on the corresponding input path shares. Note that in this study, a "path" is defined with specific boarding and

transfer stations and lines. We assume passengers following a path will only board vehicles belonging to the specific line, even though there are multiple lines that serve a trip segment. Hence, there is no "common line" problem (De Cea and Fernández, 1993) in this study because "common lines" will be treated as different paths.

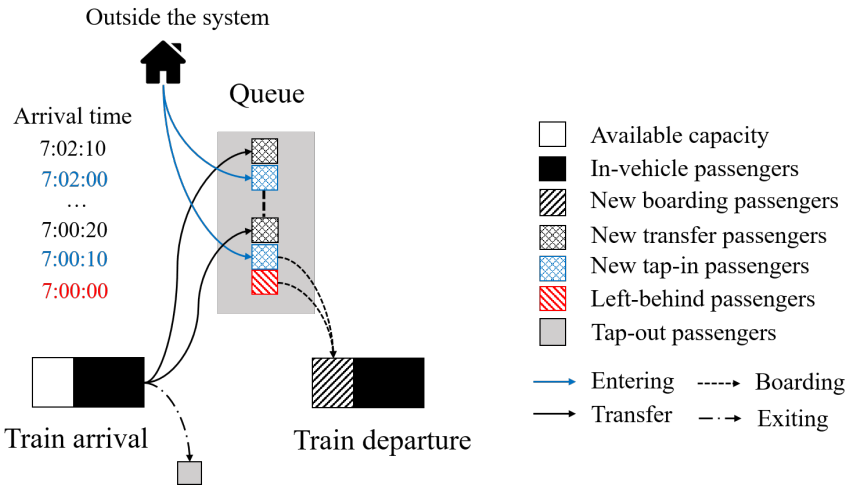


Figure 1: Structure of the network loading model (adapted from Mo et al. (2020)

For an arrival event, the train offloads passengers who reach their destination or need to transfer at the station and updates its state (e.g. train load and in-vehicle passengers). For passengers who reach their destinations, their tap-out times are calculated by adding their egress time. For those who transfer at the station, their arrival times at the next platform are calculated based on the transfer time. The transfer passengers are added to the waiting queue in order of their arrival times at the next platform.

For departure events, the queue on the platform is updated by the new tap-in passengers, that is, passengers who arrive at the platform after the last train departed are added to the queue based on their arrival times. Passengers board the train according to a First-Come-First-Serve (FCFS) discipline until the train reaches its capacity. Passengers who cannot board are left behind and wait in the queue for the next train. The states of the train and the waiting queue are updated accordingly.

The simulator can record every passenger's trajectory during the whole travel process, including tap-in time, platform arrival time, boarding time, alighting time, tap-out time, etc.

3.1.2. Simulating service disruptions

Given a service disruption, the event list is modified to incorporate the incident's impact on the supply side. Specifically, all incidents' impacts can be reflected by changes in vehicles' arrival and departure times. For example, the blockage of a rail line can be represented by some vehicles in the line having long dwell times at the corresponding stations during the incident period. The dispatching of shuttle buses can be seen as adding a new set of events (vehicle arrivals and departures) associated with the new bridging route. The headway adjustment of existing routes can also be captured by the new vehicle arrival and departure times. In this way, the event-based simulator can conveniently model service disruptions without changing the framework. It is worth noting that using the change of timetable to capture the incident impact on supply is also applicable to multi-platform scenarios (i.e., a platform serving different lines or different types of train capacities). Different from typical re-scheduling problems where the design of the new timetable needs to

consider the train conflicts in the multi-platform scenario, in this study, the timetable is given, where the possible conflicts are already considered in the new timetable. In addition, the timetable change can also capture the "partially blocked" platform. Details are illustrated in Appendix A.

From the passenger side, when an incident happens, all passengers in blocked trains are offloaded to the nearest platform. Depending on the input path choices (i.e., recommendation strategies), offloading passengers are re-assigned to a new alternative path and join the queues at the corresponding boarding station. After reassigning the offloading passengers, the simulator continues to run from the incident time to the end of the simulation period (note that passengers who have not entered the system when the incident occurs will have a new path choice depending on the input path choices).

3.2. Problem description

Consider a service disruption in an urban rail system starting at time T_s and ending at T_e. During the disruption, some stations in the incident line (or the whole line) are blocked. Passengers in the blocked trains are usually offloaded to the nearest platforms. To respond to the incident, some changes in the operations are made, such as dispatching shuttle buses, rerouting existing services, short-turning in the incident line, headway adjustment, etc. Assume that we have all information about the operating changes[1]. These changes define a new PT service network and alternative path sets. Our objective is to design an origin-destination (OD) based recommendation system. That is, when the incident happens, passengers can use their phones, websites, or electrical boards at stations to access the recommendation system. They input their **origin station, destination station, and departure time** to get a recommended path. The recommendation aims to minimize the system travel time, that is, the sum of all passengers' travel times, including passengers at nearby lines or bus routes without incidents (note that these passengers may experience additional crowding due to transfer passengers from the incident line).

Let $\mathcal{K}$ be the predetermined set of all OD pairs that may need path recommendations. $\mathcal{K}$ is defined based on whether an OD pair is affected by the incident or not. Operators usually need a period called "response time" (e.g., 10 to 20 minutes) to generate the service changes. Let the response time be η. We assume that the path recommendations start at $T_s + \eta$. Note that the origins for passengers who are already in the system at time $T_s + \eta$ (e.g., offloaded passengers from the blocked vehicles) is their current locations (as opposed to their initial origins such as the boarding stations). We aim to provide recommendations for passengers whose OD pairs are in $\mathcal{K}$ and departure times are in the range from $T_s + \eta$ to some time point after T_e, since the congestion may last longer than T_e and passengers departing after T_e may also need guidance. Suppose that the period of recommendation starts at a time point (h_0) and consists of time intervals ($h_1, ..., h_H$) of equal length τ (e.g., 10 minutes). Specifically, h_0 represents the time point at $T_s + \eta$. Recommendations at $T_s + \eta$ focus on passengers who are offloaded from blocked vehicles or arrive between T_s and $T_s + \eta$ (their departure times are $T_s + \eta$)[2]. And h_t $(t \geq 1)$ represents the time interval $(T_s + \eta + (t-1)\tau, T_s + \eta + t\tau]$. Recommendations at h_t $(t \geq 1)$ focus on passengers who were not in the system when the incident happened and their departure times are in $(T_s + \eta + (t-1)\tau, T_s + \eta + t\tau]$ (or passengers who are in the system when the incident happens but scheduled to depart in $(T_s + \eta + (t-1)\tau, T_s + \eta + t\tau]$). Let the set of all recommendation times be $\mathcal{H} := \{h_0, h_1, ..., h_H\}$. It is worth noting that, the following description aims

[1] That is, we assume during the disruption, operators would first change the supply to accommodate for the disruption, then provide path recommendations that incorporate the supply changes.

[2] Note that some of those passengers may schedule their departure times after $T_s + \eta$. These passengers will be considered as demand in other time intervals

to solve the model at time point h_0 and generate path recommendations from h_0 to h_H. However, the methodology is easy to be extended to a rolling horizon implementation where the model can be solved at any given time interval $\tilde{h} \in \mathcal{H}$. In this way, the service operation and demand information can be updated over time. Details of this discussion can be found in Section 4.1.

Given the new operations during the incident, we obtain a feasible path set R_k for each OD pair k. Note that R_k includes all feasible services that are provided by the PT operator. A path $r \in R_k$ may be waiting for the system to recover (i.e., using the incident line), or transfer to nearby bus lines, using shuttle services, etc. We do not consider non-PT modes, such as Uber or driving for the following reasons: 1) The study aims to design a path recommendation system used by PT operators to provide path alternative recommendations to all PT users. Considering non-PT modes needs the supply information of all other travel modes and even consider non-PT users (such as the impact of traffic congestion on drivers), which is beyond the scope of this study. Future research may consider a multi-modal path recommendation system. 2) Passengers using non-PT modes can be simply treated as demand reduction for the PT system. So their impact on the PT system is still captured.

Let d_{hk} be the number of passengers using the PT system with OD pair $k \in \mathcal{K}$ and departure time $h \in \mathcal{H}$. It can be treated as the normal demand minus the number of passengers leaving the PT system. As we do not have full information about future demand and the number of passengers leaving the system, d_{hk} is an uncertainty variable that will be discussed in Section 3.4. Let f_{hkr} be the number of passengers departing at time interval h using OD pair k and path $r \in R_k$. By definition:

$$\sum_{r \in R_k} f_{hkr} = d_{hk} \quad \forall h \in \mathcal{H}, k \in \mathcal{K} \tag{2}$$

Let p_{hkr} be the corresponding path share of f_{hkr} (i.e., $p_{hkr} = f_{hkr}/d_{hk}$ and $\sum_{r \in R_k} p_{hkr} = 1$). For convenience of description, we define $\mathcal{F} := \{(h, k, r) : \forall h \in \mathcal{H}, \forall k \in \mathcal{K}, r \in R_k\}$ as the set of all path indices. Then the optimal flow problem can be formulated as:

$$\min_{\boldsymbol{f}, \boldsymbol{p}} \quad Z(\boldsymbol{f}) = \text{Sum of all passengers' travel time} \tag{3a}$$

$$\text{s.t.} \quad \sum_{r \in R_k} p_{hkr} = 1 \qquad \forall\, h \in \mathcal{H}, k \in \mathcal{K}, \tag{3b}$$

$$f_{hkr} = d_{hk} \cdot p_{hkr} \quad \forall\, (h, k, r) \in \mathcal{F}, \tag{3c}$$

$$f_{hkr} \geq 0 \qquad \forall\, (h, k, r) \in \mathcal{F}, \tag{3d}$$

$$0 \leq p_{hkr} \leq 1 \qquad \forall\, (h, k, r) \in \mathcal{F} \tag{3e}$$

where $\boldsymbol{f} := (f_{hkr})_{h,k,r \in \mathcal{F}}$ and $\boldsymbol{p} := (p_{hkr})_{h,k,r \in \mathcal{F}}$. $Z(\boldsymbol{f})$ is the system travel time which has no analytical expression. It can only be obtained after each network loading or simulation process (see Section 2.4). Note that using both $\boldsymbol{f}$ and $\boldsymbol{p}$ in the optimization problem is redundant, but it is useful for explaining the methodology.

If there is no uncertainty in the system, the optimal path shares (p^*_{hkr}) obtained from the solution of Eq. 3 are the recommendation proportions. That is, for all passengers with OD pair k and departure time h, the system will recommend them to use path r with probability p^*_{hkr}. However, Eq. 3 is a conceptual formulation, it cannot be solved directly because $Z(\boldsymbol{f})$ has no analytical expression. Moreover, given the uncertainties in demand, the final recommended path shares may not be p^*_{hkr}. In the following sections, we

elaborate on how to solve the robust "optimal flow problem" with demand uncertainties.

3.3. Simulation-based linearization of the objective function

In this section, we propose a simulation-based linearization for the non-analytical $Z(\boldsymbol{f})$ based on a first-order approximation. $Z(\boldsymbol{f})$ can be approximated as:

$$\hat{Z}(\boldsymbol{f}) = Z(\tilde{\boldsymbol{f}}) + (\boldsymbol{f} - \tilde{\boldsymbol{f}})^T \frac{\partial Z(\boldsymbol{f})}{\partial \boldsymbol{f}}|_{\boldsymbol{f}=\tilde{\boldsymbol{f}}} \tag{4}$$

where $\hat{Z}(\boldsymbol{f})$ is the first-order approximation of $Z(\boldsymbol{f})$. $\tilde{\boldsymbol{f}}$ is a reference flow for the first-order approximation. $Z(\tilde{\boldsymbol{f}})$ is the system travel time estimated by simulation with $\tilde{\boldsymbol{f}}$ as input. $\frac{\partial Z(\boldsymbol{f})}{\partial \boldsymbol{f}} = (\frac{\partial Z(\boldsymbol{f})}{\partial f_{hkr}})_{h,k,r\in\mathcal{F}}$ is the gradient vector of $Z(\boldsymbol{f})$. As $\tilde{\boldsymbol{f}}$ and $Z(\tilde{\boldsymbol{f}})$ are pre-determined, the only unknown part is $\frac{\partial Z(\boldsymbol{f})}{\partial \boldsymbol{f}}|_{\boldsymbol{f}=\tilde{\boldsymbol{f}}}$. Notice that $\frac{\partial Z(\boldsymbol{f})}{\partial f_{hkr}}|_{\boldsymbol{f}=\tilde{\boldsymbol{f}}}$ represents the change of system travel time caused by one unit of flow change in f_{hkr}. It can be approximated as:

$$\frac{\partial Z(\boldsymbol{f})}{\partial f_{hkr}}|_{\boldsymbol{f}=\tilde{\boldsymbol{f}}} \approx \frac{Z(\tilde{\boldsymbol{f}} + \boldsymbol{e}_{hkr}) - Z(\tilde{\boldsymbol{f}})}{1} \tag{5}$$

where $\boldsymbol{e}_{hkr}$ represents a vector with only the (h,k,r)-th element being 1 and others zero. Eq. 5 represents the numerical approximation of the gradient. Now we only need to calculate $Z(\tilde{\boldsymbol{f}} + \boldsymbol{e}_{hkr}) - Z(\tilde{\boldsymbol{f}})$. A naive method to do that is to run a simulation with $\tilde{\boldsymbol{f}} + \boldsymbol{e}_{hkr}$ as input. However, as running the simulation is time-consuming, this method is not efficient. Note that since we already run a simulation with $\tilde{\boldsymbol{f}}$ as input, it is possible to directly calculate the marginal change due to the additional unit of flow (i.e., calculate the additional travel time increase to the system if one additional flow is added to $\tilde{f}_{hkr}$).

Consider an example journey of $\tilde{f}_{hkr}$ in Figure 2. Let $\mathcal{M}_{hkr}$ be the set of passengers composing the flow of $\tilde{f}_{hkr}$ (i.e., the green passengers in Figure 2). These passengers have origin station a_1 and destination station a_7, and the path includes a transfer from station a_4 to station a_5. Let the average travel time of $\tilde{f}_{hkr}$ be $T^{\mathrm{A}}_{hkr}(\tilde{\boldsymbol{f}})$. Suppose that one more passenger is added to $\tilde{f}_{hkr}$.

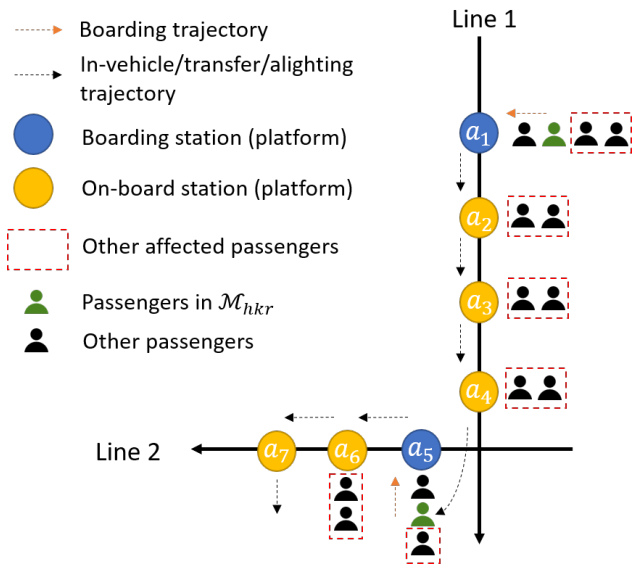


Figure 2: Explanation for the impact of adding an additional one unit flow to the system

First of all, the system travel time is increased by $T^{\mathrm{A}}_{hkr}(\tilde{\boldsymbol{f}})$ due to the increase in the flow amount. Note that considering the marginal calculation, we ignore the impact of the added passenger on the increase in $T^{\mathrm{A}}_{hkr}(\tilde{\boldsymbol{f}})$. Besides, all passengers in the red-dashed square may experience higher travel times. Passengers at station a_1 and a_5 who queue behind the green passenger may have additional waiting time if the train that $\mathcal{M}_{hkr}$ used is full after departure (under the simulation results of $\tilde{\boldsymbol{f}}$), because the increase of the flow by one in $\tilde{f}_{hkr}$ will occupy one available capacity for these waiting passengers, and one of them will have to board the next train (i.e., wait for one more headway). Denote the total increase in system travel time for passengers queuing behind $\mathcal{M}_{hkr}$ as $T^{\mathrm{Q}}_{hkr}(\tilde{\boldsymbol{f}})$. The detailed calculation of $T^{\mathrm{Q}}_{hkr}(\tilde{\boldsymbol{f}})$ is shown in Appendix B.1.

For passengers waiting at stations where $\mathcal{M}_{hkr}$ are already on-board (referred to as on-board stations, e.g., station a_2), adding one flow to $\tilde{f}_{hkr}$ reduces the available capacity when the vehicle arrives at these on-board stations. The queuing passengers at the onboard stations may not be able to board due to the reduction of capacity. Specifically, if a vehicle is full when it departs from an onboard station under flow pattern $\tilde{\boldsymbol{f}}$, adding one passenger to $\tilde{f}_{hkr}$ makes one passenger waiting at the on-board station unable to board his/her original boarded vehicle. And the system travel time is increased by one headway for each of these onboard stations. Denote the travel time increase for passengers waiting at on-board stations as $T^{\mathrm{O}}_{hkr}(\tilde{\boldsymbol{f}})$. The detailed calculation of $T^{\mathrm{O}}_{hkr}(\tilde{\boldsymbol{f}})$ is shown in Appendix B.2.

Therefore, in this way, depending on whether the vehicle is full or not under flow pattern $\tilde{\boldsymbol{f}}$, the increase in system travel time due to adding one passenger to $\tilde{f}_{hkr}$ can be calculated without running the simulation again. These increases come from three parts: 1) the average travel time of $\mathcal{M}_{hkr}$ due to increasing in flow amount (i.e., $T^{\mathrm{A}}_{hkr}(\tilde{\boldsymbol{f}})$), 2) the additional waiting time for passengers queuing behind $\mathcal{M}_{hkr}$ (i.e., $T^{\mathrm{Q}}_{hkr}(\tilde{\boldsymbol{f}})$), and 3) the additional waiting time for passengers queuing at $\mathcal{M}_{hkr}$'s on-board stations (i.e., $T^{\mathrm{O}}_{hkr}(\tilde{\boldsymbol{f}})$). Specifically, we have

$$Z(\tilde{\boldsymbol{f}} + \boldsymbol{e}_{hkr}) - Z(\tilde{\boldsymbol{f}}) = T^{\mathrm{A}}_{hkr}(\tilde{\boldsymbol{f}}) + T^{\mathrm{Q}}_{hkr}(\tilde{\boldsymbol{f}}) + T^{\mathrm{O}}_{hkr}(\tilde{\boldsymbol{f}}) \tag{6}$$

Consequently, $\frac{\partial Z(\boldsymbol{f})}{\partial \boldsymbol{f}}|_{\boldsymbol{f}=\tilde{\boldsymbol{f}}}$ can be obtained from Eq. 5. Define $\boldsymbol{\beta}(\tilde{\boldsymbol{f}}) := \frac{\partial Z(\boldsymbol{f})}{\partial \boldsymbol{f}}|_{\boldsymbol{f}=\tilde{\boldsymbol{f}}}$. Then the objective function becomes:

$$\hat{Z}(\boldsymbol{f}) = Z(\tilde{\boldsymbol{f}}) + \boldsymbol{\beta}(\tilde{\boldsymbol{f}})^T(\boldsymbol{f} - \tilde{\boldsymbol{f}}) \tag{7}$$

where $\boldsymbol{\beta}(\tilde{\boldsymbol{f}}) = (\beta_{hkr})_{h,k,r\in\mathcal{F}}$ and $\beta_{hkr} = \frac{\partial Z(\boldsymbol{f})}{\partial f_{hkr}}|_{\boldsymbol{f}=\tilde{\boldsymbol{f}}}$. Eq. 7 is a linear function of $\boldsymbol{f}$, which supports for addressing uncertainties in the optimization problem.

3.4. Demand uncertainty

The uncertainty of d_{hk} comes from two different parts. The first is the inherent demand variations across different days, and the second is the uncertainty in how many passengers leave the PT system during the incident. In this section, these two uncertainties are considered as a whole by introducing an ellipsoidal uncertainty set and three polyhedral uncertainty sets.

From constraint 3c, we can substitute $f_{hkr} = d_{hk} \cdot p_{hkr}$ to the objective function and rewrite Eq. 7 as:

$$\hat{Z}(\boldsymbol{f}) = \hat{Z}(\boldsymbol{p}) = Z(\tilde{\boldsymbol{f}}) + \sum_{(h,k,r)\in\mathcal{F}} \beta_{hkr} \cdot (d_{hk} \cdot p_{hkr} - \tilde{f}_{hkr}) \tag{8}$$

Note that β_{hkr} is a function of $\tilde{\boldsymbol{f}}$, for simplicity we ignore $\tilde{\boldsymbol{f}}$ in the derivation process.

To model the uncertainty of d_{hk}, we introduce an auxiliary decision variable t and rewrite the optimal flow problem as:

$$\min_{\boldsymbol{p}, t} \quad t \tag{9a}$$

$$\text{s.t.} \quad t \geq Z(\tilde{\boldsymbol{f}}) + \sum_{(h,k,r)\in\mathcal{F}} \beta_{hkr} \cdot (d_{hk} \cdot p_{hkr} - \tilde{f}_{hkr}), \tag{9b}$$

$$\text{Constraints (3b) and (3e)} \tag{9c}$$

Constraint 9b can be rewritten as

$$\sum_{h,k} \sum_{r\in R_k} \beta_{hkr} \cdot d_{hk} \cdot p_{hkr} \leq t - Z(\tilde{\boldsymbol{f}}) + \sum_{(h,k,r)\in\mathcal{F}} \beta_{hkr} \tilde{f}_{hkr} \tag{10}$$

Eq. 10 can be written in a matrix form as:

$$\boldsymbol{a}^T \boldsymbol{p} \leq b \tag{11}$$

where $\boldsymbol{a} \in \mathbb{R}^{|\mathcal{F}|}$ with the entry $a_{hkr} = \beta_{hkr} d_{hk},\ \forall\ (h,k,r) \in \mathcal{F}$. And $b = t - Z(\tilde{\boldsymbol{f}}) + \sum_{(h,k,r)\in\mathcal{F}} \beta_{hkr} \tilde{f}_{hkr}$. Define $\boldsymbol{d} = (d_{hk})_{h\in\mathcal{H}, k\in\mathcal{K}}$.

Proposition 1. *If $\boldsymbol{d}$ is normally distributed with $\boldsymbol{d} \sim \mathcal{N}(\bar{\boldsymbol{d}}, \boldsymbol{\Sigma})$, then in a RO problem where constraint 11 is guaranteed to be satisfied with probability of at least $1-\varepsilon$ (i.e., $\mathbb{P}[\boldsymbol{a}^T\boldsymbol{p} \leq b] \geq 1-\varepsilon$), the robust constraint can be formulated as:*

$$(\boldsymbol{A}\bar{\boldsymbol{d}} + \boldsymbol{A}\boldsymbol{D}\boldsymbol{z})^T \boldsymbol{p} \leq b, \quad \forall \boldsymbol{z} \in \mathcal{Z}_E \tag{12}$$

where $\boldsymbol{A} \in \mathbb{R}^{|\mathcal{F}|\times HK}$ with entry $A_{hkr,h'k'} = \beta_{hkr}$ if $h = h'$ and $k = k'$, otherwise $A_{hkr,h'k'} = 0$. $\boldsymbol{D}$ is the Cholesky decomposition of $\boldsymbol{\Sigma}$ (i.e., $\boldsymbol{\Sigma} = \boldsymbol{D}\boldsymbol{D}^T$). $\boldsymbol{z}$ are the perturbation variables (i.e., $\boldsymbol{d} = \bar{\boldsymbol{d}} + \boldsymbol{D}\boldsymbol{z}$) and $\mathcal{Z}_E = \left\{\boldsymbol{z} \in \mathbb{R}^{HK} : \|z\|_2 \leq \rho_{1-\varepsilon}\right\}$ (i.e., the ellipsoidal uncertainty set). $\rho_{1-\varepsilon}$ is the $(1-\varepsilon)$-percentile of a standard normal distribution.

Proof.

Step 1: We first prove that $\mathbb{P}[\boldsymbol{a}^T\boldsymbol{p} \leq b] \geq 1-\varepsilon$ is equivalent to $(\boldsymbol{A}\bar{\boldsymbol{d}})^T\boldsymbol{p} + \rho_{1-\varepsilon} \left\|(\boldsymbol{A}\boldsymbol{D})^T\boldsymbol{p}\right\|_2 \leq b$.

Since $\boldsymbol{d}$ is normally distributed, we have $\boldsymbol{a} = \boldsymbol{A}\boldsymbol{d}$ is normally distributed with $\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{A}\bar{\boldsymbol{d}}, \boldsymbol{A}\Sigma\boldsymbol{A}^T)$. Similarly, $\boldsymbol{a}^T\boldsymbol{p} \in \mathbb{R}$ is also normally distributed with

$$\boldsymbol{a}^T\boldsymbol{p} \sim \mathcal{N}((\boldsymbol{A}\bar{\boldsymbol{d}})^T\boldsymbol{p}, \boldsymbol{p}^T\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T\boldsymbol{p}) \tag{13}$$

If we want constraint 11 to hold with probability at least $1-\varepsilon$, it suffices to have:

$$(\boldsymbol{A}\bar{\boldsymbol{d}})^T\boldsymbol{p} + \rho_{1-\varepsilon}\sqrt{\boldsymbol{p}^T\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T\boldsymbol{p}} \leq b \tag{14}$$

Substituting $\boldsymbol{\Sigma} = \boldsymbol{D}\boldsymbol{D}^T$ into Eq. 14 completes the proof of Step 1.

Step 2: We need to show that the robust counterpart of Eq. 12 is $(\boldsymbol{A}\bar{\boldsymbol{d}})^T\boldsymbol{p} + \rho_{1-\varepsilon} \left\|(\boldsymbol{A}\boldsymbol{D})^T\boldsymbol{p}\right\|_2 \leq b$.

Eq. 12 is equivalent to:

$$(\boldsymbol{A}\bar{\boldsymbol{d}})^T\boldsymbol{p} + \max_{\boldsymbol{z}\in\mathcal{Z}_{\mathrm{E}}}(\boldsymbol{A}\boldsymbol{D}\boldsymbol{z})^T\boldsymbol{p} \leq b. \tag{15}$$

Let $\delta(\boldsymbol{z} \mid \mathcal{Z}_{\mathrm{E}})$ be the indicator function on set $\mathcal{Z}_{\mathrm{E}}$:

$$\delta(\boldsymbol{z} \mid \mathcal{Z}_{\mathrm{E}}) = \begin{cases} 1, & \text{if } \boldsymbol{z} \in \mathcal{Z}_{\mathrm{E}} \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

Then the convex conjugate of $\delta(\boldsymbol{z} \mid \mathcal{Z}_{\mathrm{E}})$ (also known as the **support function**) can be derived as (Bertsimas and den Hertog, 2020):

$$\delta^*(\boldsymbol{y} \mid \mathcal{Z}_{\mathrm{E}}) = \sup_{\boldsymbol{z}\in\mathbb{R}^{HK}}\{\boldsymbol{y}^T\boldsymbol{z} - \delta(\boldsymbol{z} \mid \mathcal{Z}_{\mathrm{E}})\} = \sup_{\boldsymbol{z}\in\mathcal{Z}_{\mathrm{E}}} \boldsymbol{y}^T\boldsymbol{z} = \rho_{1-\varepsilon}\left\|\boldsymbol{y}\right\|_2 \tag{17}$$

Therefore, Eq. 15 can be rewritten with the convex conjugate:

$$(\boldsymbol{A}\bar{\boldsymbol{d}})^T\boldsymbol{p} + \delta^*((\boldsymbol{A}\boldsymbol{D})^T\boldsymbol{p} \mid \mathcal{Z}) = (\boldsymbol{A}\bar{\boldsymbol{d}})^T\boldsymbol{p} + \rho_{1-\varepsilon}\left\|(\boldsymbol{A}\boldsymbol{D})^T\boldsymbol{p}\right\|_2 \leq b \tag{18}$$

which finishes the proof of Step 2. Combining Steps 1 and 2 finishes the proof of the whole proposition. □

We observe that the ellipsoidal demand uncertainty performs like a regularization. It prevents $\boldsymbol{p}$ from being large in directions with considerable uncertainty in the demand.

Remark 1. In the RO, the ellipsoidal uncertainty set can be used no matter what distribution $\boldsymbol{d}$ follows. If $\boldsymbol{d}$ is normally distributed, the parameter $\rho_{1-\varepsilon}$ can be interpreted as the probability that constraint 11 holds. The use of the multivariate normality assumption in Proposition 1 is for explaining the physical meaning of ellipsoidal uncertainty set and facilitating the choice of hyperparameters (i.e., $\rho_{1-\varepsilon}$ and $\boldsymbol{D}$). Moreover, in the case study, we partially validate the multivariate normality assumption of $\boldsymbol{d}$ using smart card data. The Mardia's Skewness Test (Cain et al., 2017) shows that $\boldsymbol{d}$ has no significant skewness.

Eq. 12 (i.e., the ellipsoidal uncertainty set) captures the correlation between demands at different time intervals and OD pairs. However, it does not impose any upper or lower bounds on d_{hk}. In reality, the demand level for a specific OD pair and time interval is usually bounded, which can be expressed as:

$$d_{hk}^{\mathrm{L}} \leq d_{hk} \leq d_{hk}^{\mathrm{U}} \tag{19}$$

where d_{hk}^{L} and d_{hk}^{U} are the corresponding lower and upper bounds for d_{hk}, respectively. Their values can be obtained from historical demand data. Eq. 19 can be rewritten in a vector form as $\boldsymbol{d}^{\mathrm{L}} \leq \boldsymbol{d} \leq \boldsymbol{d}^{\mathrm{U}}$, where $\boldsymbol{d}^{\mathrm{U}} = (d_{hk}^{\mathrm{U}})_{h\in\mathcal{H},k\in\mathcal{K}}$ and $\boldsymbol{d}^{\mathrm{L}} = (d_{hk}^{\mathrm{L}})_{h\in\mathcal{H},k\in\mathcal{K}}$. Since we have $\boldsymbol{d} = \bar{\boldsymbol{d}} + \boldsymbol{D}\boldsymbol{z}$, a simple manipulation leads to

$$\boldsymbol{d}^{\mathrm{L}} - \bar{\boldsymbol{d}} \leq \boldsymbol{D}\boldsymbol{z} \leq \boldsymbol{d}^{\mathrm{U}} - \bar{\boldsymbol{d}} \tag{20}$$

We can rewrite it as a "polyhedral uncertainty set": $\mathcal{Z}_{\mathrm{P1}} = \left\{\boldsymbol{z} \in \mathbb{R}^{HK} : \boldsymbol{d}^{\mathrm{L}} - \bar{\boldsymbol{d}} \leq \boldsymbol{D}\boldsymbol{z} \leq \boldsymbol{d}^{\mathrm{U}} - \bar{\boldsymbol{d}}\right\}$.

Eq. 19 ensures the boundaries for each individual demand. Another similar constraint for the demand uncertainty is that: within a given time interval, the total demand across all OD pairs should also be bounded. This constraint can avoid some extreme scenarios that Eq. 19 cannot capture (e.g., all d_{hk} are at the lower

or upper bounds). Mathematically:

$$d_h^{\mathrm{L}} \leq \sum_{k \in \mathcal{K}} d_{hk} \leq d_h^{\mathrm{U}} \tag{21}$$

where d_h^{L} and d_h^{U} are the lower and upper bounds for the total demand in time interval h, which can be obtained from the historical demand. Define $\boldsymbol{S} \in \mathbb{R}^{H \times HK}$, where the element $S_{h,h'k} = 1$ if $h = h'$, otherwise $S_{h,h'k} = 0$. Then Eq. 21 can be rewritten in a matrix form:

$$\boldsymbol{d}_{\mathcal{H}}^{\mathrm{L}} - \boldsymbol{S}\bar{\boldsymbol{d}} \leq \boldsymbol{S}\boldsymbol{D}\boldsymbol{z} \leq \boldsymbol{d}_{\mathcal{H}}^{\mathrm{U}} - \boldsymbol{S}\bar{\boldsymbol{d}} \tag{22}$$

where $\boldsymbol{d}_{\mathcal{H}}^{\mathrm{U}} = (d_h^{\mathrm{U}})_{h \in \mathcal{H}}$ and $\boldsymbol{d}_{\mathcal{H}}^{\mathrm{L}} = (d_h^{\mathrm{L}})_{h \in \mathcal{H}}$. And Eq. 22 can also be represented as a polyhedral uncertainty set: $\mathcal{Z}_{\mathrm{P2}} = \left\{\boldsymbol{z} \in \mathbb{R}^{HK} : \boldsymbol{d}_{\mathcal{H}}^{\mathrm{L}} - \boldsymbol{S}\bar{\boldsymbol{d}} \leq \boldsymbol{S}\boldsymbol{D}\boldsymbol{z} \leq \boldsymbol{d}_{\mathcal{H}}^{\mathrm{U}} - \boldsymbol{S}\bar{\boldsymbol{d}}\right\}$.

As the RO aims to optimize under the "worst case" scenario and our objective function is the system travel time, intuitively, the worst-case scenario will be the largest demand in the uncertainty set. This may make the worst-case demand unrealistic since the extremely large demand seldom happens. What we expect in the RO is that the model can capture some critical OD pairs where the high demand in these OD pairs can make the system more congested (as opposed to high demand in all OD pairs). In order to let the RO capture critical OD pairs, we add an additional constraint on the total demand:

$$\sum_{h \in \mathcal{H}, k \in \mathcal{K}} d_{hk} \leq \Gamma \cdot \sum_{h \in \mathcal{H}, k \in \mathcal{K}} \bar{d}_{hk} \tag{23}$$

where $\Gamma > 0$ is a predetermined constant. $\Gamma = 1$ means we assume the total demand in the worst-case scenario is the same as the nominal one, but the spatial and temporal distributions are different. The worst-case scenario will have more demand on critical OD pairs but less demand on others. The value of Γ can be determined based on the highest total demand observed over a time period.

Similarly, Eq. 23 can be written in a matrix form:

$$\mathbf{1}^T(\bar{\boldsymbol{d}} + \boldsymbol{D}\boldsymbol{z}) \leq \Gamma \cdot \mathbf{1}^T\bar{\boldsymbol{d}} \tag{24}$$

where $\mathbf{1} \in \mathbb{R}^{HK}$ is a vector with all elements one. And we define another polyhedral uncertainty set: $\mathcal{Z}_{\mathrm{P3}} = \left\{\boldsymbol{z} \in \mathbb{R}^{HK} : \mathbf{1}^T(\bar{\boldsymbol{d}} + \boldsymbol{D}\boldsymbol{z}) \leq \Gamma \cdot \mathbf{1}^T\bar{\boldsymbol{d}}\right\}$.

Therefore, the final robust constraint for Eq. 11 is

$$(\boldsymbol{A}\bar{\boldsymbol{d}} + \boldsymbol{A}\boldsymbol{D}\boldsymbol{z})^T\boldsymbol{p} \leq b, \quad \forall \boldsymbol{z} \in \mathcal{Z}_{\mathrm{E}} \cap \mathcal{Z}_{\mathrm{P}} \cap \mathcal{Z}_{\mathrm{P2}} \cap \mathcal{Z}_{\mathrm{P3}} \tag{25}$$

To derive the robust counterpart of the constraint, we first introduce the following lemma.

Lemma 1. *For a constraint $\bar{\boldsymbol{a}}^T\boldsymbol{x} + \delta^*(\boldsymbol{P}^T\boldsymbol{x} \mid \mathcal{Z}) \leq b$, let $\mathcal{Z}_1, ..., \mathcal{Z}_k$ be closed convex sets, such that $\bigcap_i ri(\mathcal{Z}_i) \neq \emptyset$[3], and let $\mathcal{Z} = \cap_{i=1}^k \mathcal{Z}_i$. Then,*

$$\delta^*(\boldsymbol{y} \mid \mathcal{Z}) = \min_{\boldsymbol{y}_1, ..., \boldsymbol{y}_k} \{\sum_{i=1}^k \delta^*(\boldsymbol{y}_i \mid \mathcal{Z}_i) \mid \sum_{i=1}^k \boldsymbol{y}_i = \boldsymbol{y}\},$$

[3] $ri(\mathcal{Z}_i)$ indicates the relative interior of the set $\mathcal{Z}_i$.

and the constraint becomes

$$\begin{cases} \bar{\boldsymbol{a}}^T \boldsymbol{x} + \sum_{i=1}^k \delta^*(\boldsymbol{y}_i \mid \mathcal{Z}_i) \le b \\ \sum_{i=1}^k \boldsymbol{y}_i = \boldsymbol{P}^T \boldsymbol{x} \end{cases}$$

where $\delta^*(\cdot \mid \cdot)$ is the support function (i.e., convex conjugate of the indicator function).

The proof of Lemma 1 can be found in Ben-Tal et al. (2015). From Proposition 1, we have $\delta^*(\boldsymbol{y} \mid \mathcal{Z}_{\text{E}}) = \rho_{1-\varepsilon} \|\boldsymbol{y}\|_2$. For the polyhedral uncertainty set, consider a general form $\mathcal{Z}_{\text{P}} = \{\boldsymbol{z} : \boldsymbol{H}\boldsymbol{z} \le \boldsymbol{c}\}$. And the support function for $\mathcal{Z}_{\text{P}}$ is

$$\delta^*(\boldsymbol{y} \mid \mathcal{Z}_{\text{P}}) = \max_{\boldsymbol{z}} \{\boldsymbol{y}^T \boldsymbol{z} \mid \boldsymbol{H}\boldsymbol{z} \le \boldsymbol{c}\} = \min_{\boldsymbol{u}} \{\boldsymbol{c}^T \boldsymbol{u} \mid \boldsymbol{H}^T \boldsymbol{u} = \boldsymbol{y}, \boldsymbol{u} \ge 0\} \tag{26}$$

where the second equality follows from linear programming duality. Eq. 26 can be used to derive the support function for $\mathcal{Z}_{\text{P1}}$, $\mathcal{Z}_{\text{P2}}$, and $\mathcal{Z}_{\text{P3}}$. For example, consider the robust counterpart for Eq. 23, we have

$$\delta^*(\boldsymbol{y}_6 \mid \mathcal{Z}_{\text{P3}}) = \min_{u_3} \{(\Gamma - 1) \cdot (\boldsymbol{1}^T \bar{\boldsymbol{d}}) \cdot u_3 \mid (\boldsymbol{1}^T \boldsymbol{D})^T u_3 = \boldsymbol{y}_6, u_3 \ge 0\} \tag{27}$$

where $\boldsymbol{y}_6 \in \mathbb{R}^{HK}$ and $u_3 \in \mathbb{R}$ are decision variables in the RO model. Note that the subscripts for $\boldsymbol{y}$ and u (i.e., 6 and 3) are used for the consistency in Eq. 28.

Based on Lemma 1, the robust counterpart for Eq. 25 is

$$\begin{aligned}
&(\boldsymbol{A}\bar{\boldsymbol{d}})^T \boldsymbol{p} + \rho_{1-\varepsilon} \|\boldsymbol{y}_1\|_2 + (\boldsymbol{d}^{\text{U}} - \bar{\boldsymbol{d}})^T \boldsymbol{u}_1 + (\bar{\boldsymbol{d}} - \boldsymbol{d}^{\text{L}})^T \boldsymbol{u}_2 + (\boldsymbol{d}^{\text{U}}_{\mathcal{H}} - \boldsymbol{S}\bar{\boldsymbol{d}})^T \boldsymbol{v}_1 + (\boldsymbol{S}\bar{\boldsymbol{d}} - \boldsymbol{d}^{\text{L}}_{\mathcal{H}})^T \boldsymbol{v}_2 \\
&+ (\Gamma - 1) \cdot (\boldsymbol{1}^T \bar{\boldsymbol{d}}) \cdot u_3 \le b && \text{(28a)} \\
&\boldsymbol{D}^T \boldsymbol{u}_1 = \boldsymbol{y}_2 && \text{(28b)} \\
&-\boldsymbol{D}^T \boldsymbol{u}_2 = \boldsymbol{y}_3 && \text{(28c)} \\
&(\boldsymbol{S}\boldsymbol{D})^T \boldsymbol{v}_1 = \boldsymbol{y}_4 && \text{(28d)} \\
&-(\boldsymbol{S}\boldsymbol{D})^T \boldsymbol{v}_2 = \boldsymbol{y}_5 && \text{(28e)} \\
&(\boldsymbol{1}^T \boldsymbol{D})^T u_3 = \boldsymbol{y}_6 && \text{(28f)} \\
&\sum_{i=1}^{6} \boldsymbol{y}_i = (\boldsymbol{A}\boldsymbol{D})^T \boldsymbol{p} && \text{(28g)} \\
&\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{v}_1, \boldsymbol{v}_2, u_3 \ge 0 && \text{(28h)}
\end{aligned}$$

Hence, the RO problem can be formulated as

$$\begin{aligned}
\min_{\boldsymbol{p}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{y}, t} \quad & t && \text{(29a)} \\
\text{s.t.} \quad & \sum_{(h,k,r) \in \mathcal{F}} \beta_{hkr} \cdot d_{hk} \cdot p_{hkr} + \rho_{1-\varepsilon} \|\boldsymbol{y}_1\|_2 + (\boldsymbol{d}^{\text{U}} - \bar{\boldsymbol{d}})^T \boldsymbol{u}_1 + (\bar{\boldsymbol{d}} - \boldsymbol{d}^{\text{L}})^T \boldsymbol{u}_2 + (\boldsymbol{d}^{\text{U}}_{\mathcal{H}} - \boldsymbol{S}\bar{\boldsymbol{d}})^T \boldsymbol{v}_1 \\
& + (\boldsymbol{S}\bar{\boldsymbol{d}} - \boldsymbol{d}^{\text{L}}_{\mathcal{H}})^T \boldsymbol{v}_2 + (\Gamma - 1) \cdot (\boldsymbol{1}^T \bar{\boldsymbol{d}}) \cdot u_3 + Z(\tilde{\boldsymbol{f}}) - \sum_{(h,k,r) \in \mathcal{F}} \beta_{hkr} \tilde{f}_{hkr} \le t && \text{(29b)} \\
& \text{Constraints } (28b) - (28h) && \text{(29c)} \\
& \text{Constraints } (3b) \text{ and } (3e) && \text{(29d)}
\end{aligned}$$

By eliminating t and inserting constraint 29b in the objective function it becomes

$$\hat{Z}(\boldsymbol{p},\boldsymbol{u},\boldsymbol{v},\boldsymbol{y})^{\mathrm{RC}} = \sum_{(h,k,r)\in\mathcal{F}} \beta_{hkr}\cdot(d_{hk}\cdot p_{hkr} - \tilde{f}_{hkr}) + \rho_{1-\varepsilon}\left\|\boldsymbol{y}_1\right\|_2 + (\boldsymbol{d}^{\mathrm{U}} - \bar{\boldsymbol{d}})^T\boldsymbol{u}_1 + (\bar{\boldsymbol{d}} - \boldsymbol{d}^{\mathrm{L}})^T\boldsymbol{u}_2 + (\boldsymbol{d}^{\mathrm{U}}_{\mathcal{H}} - \boldsymbol{S}\bar{\boldsymbol{d}})^T\boldsymbol{v}_1 + (\boldsymbol{S}\bar{\boldsymbol{d}} - \boldsymbol{d}^{\mathrm{L}}_{\mathcal{H}})^T\boldsymbol{v}_2 + (\Gamma-1)\cdot(\boldsymbol{1}^T\bar{\boldsymbol{d}})\cdot u_3 + Z(\tilde{\boldsymbol{f}}) \tag{30}$$

which yields a second-order cone programming (SOCP).

3.5. Solution procedure

After incorporating the demand uncertainty, the final robust counterpart (RC) of the optimal flow problem can be formulated as:

$$[RC(\tilde{\boldsymbol{f}})] \quad \min_{\boldsymbol{p},\boldsymbol{u},\boldsymbol{v},\boldsymbol{y}} \quad \hat{Z}(\boldsymbol{p},\boldsymbol{u},\boldsymbol{v},\boldsymbol{y})^{\mathrm{RC}} = \sum_{(h,k,r)\in\mathcal{F}} \beta_{hkr}(\tilde{\boldsymbol{f}})\cdot(d_{hk}\cdot p_{hkr} - \tilde{f}_{hkr}) + \rho_{1-\varepsilon}\left\|\boldsymbol{y}_1\right\|_2 + (\boldsymbol{d}^{\mathrm{U}} - \bar{\boldsymbol{d}})^T\boldsymbol{u}_1 + (\bar{\boldsymbol{d}} - \boldsymbol{d}^{\mathrm{L}})^T\boldsymbol{u}_2 + (\boldsymbol{d}^{\mathrm{U}}_{\mathcal{H}} - \boldsymbol{S}\bar{\boldsymbol{d}})^T\boldsymbol{v}_1 + (\boldsymbol{S}\bar{\boldsymbol{d}} - \boldsymbol{d}^{\mathrm{L}}_{\mathcal{H}})^T\boldsymbol{v}_2 + (\Gamma-1)\cdot(\boldsymbol{1}^T\bar{\boldsymbol{d}})\cdot u_3 + Z(\tilde{\boldsymbol{f}}) \tag{31a}$$

$$\text{s.t.} \quad \text{Constraints } (28b) - (28h) \tag{31b}$$

$$\sum_{r\in R_k} p_{hkr} = 1 \quad \forall h\in\mathcal{H}, k\in\mathcal{K} \tag{31c}$$

$$0 \le p_{hkr} \le 1 \quad \forall (h,k,r)\in\mathcal{F} \tag{31d}$$

This SOCP can be efficiently solved by inner interior point methods that are embedded in many existing solvers.

However, due to the first-order approximation of $Z(\boldsymbol{f})$, $\beta_{hkr}(\tilde{\boldsymbol{f}})$ needs to be updated once a new flow pattern is obtained. Hence, after obtaining $\boldsymbol{p}^*$ from the RC problem, the simulation should be run again to update $\beta_{hkr}(\tilde{\boldsymbol{f}})$. Before that, the corresponding worst-case demand (WD), which will be used as the new $\tilde{\boldsymbol{f}}$, is needed. It can be obtained by solving the worst case $\boldsymbol{z}\in\mathcal{Z}_{\mathrm{E}}\cap\mathcal{Z}_{\mathrm{P1}}\cap\mathcal{Z}_{\mathrm{P2}}\cap\mathcal{Z}_{\mathrm{P3}}$:

$$[WD(\boldsymbol{p}^*)] \quad \max_{\boldsymbol{z}} \quad (\boldsymbol{A}\boldsymbol{D}\boldsymbol{z})^T\boldsymbol{p}^* \tag{32a}$$

$$\text{s.t.} \quad \left\|\boldsymbol{z}\right\|_2 \le \rho_{1-\varepsilon} \tag{32b}$$

$$\boldsymbol{d}^{\mathrm{L}} - \bar{\boldsymbol{d}} \le \boldsymbol{D}\boldsymbol{z} \le \boldsymbol{d}^{\mathrm{U}} - \bar{\boldsymbol{d}} \tag{32c}$$

$$\boldsymbol{d}^{\mathrm{L}}_{\mathcal{H}} - \boldsymbol{S}\bar{\boldsymbol{d}} \le \boldsymbol{S}\boldsymbol{D}\boldsymbol{z} \le \boldsymbol{d}^{\mathrm{U}}_{\mathcal{H}} - \boldsymbol{S}\bar{\boldsymbol{d}} \tag{32d}$$

$$\boldsymbol{1}^T(\bar{\boldsymbol{d}} + \boldsymbol{D}\boldsymbol{z}) \le \Gamma\cdot\boldsymbol{1}^T\bar{\boldsymbol{d}} \tag{32e}$$

If the solution for Eq. 32 is $\boldsymbol{z}^*$, the worse case demand is $\boldsymbol{d}^* = \bar{\boldsymbol{d}} + \boldsymbol{D}\boldsymbol{z}^*$. Next, we can update $\boldsymbol{\beta}(\tilde{\boldsymbol{f}})$ and $Z(\tilde{\boldsymbol{f}})$ as

$$Z(\tilde{\boldsymbol{f}}), \boldsymbol{\beta}(\tilde{\boldsymbol{f}}) = \textsc{Sim-FOA}(\boldsymbol{d}^*, \boldsymbol{p}^*) \tag{33}$$

where $\tilde{\boldsymbol{f}}$ in Eq. 33 indicates $\tilde{f}_{hkr} = d^*_{hk}\cdot p^*_{hkr}$. And $\textsc{Sim-FOA}(\cdot)$ is a pseudo function of simulation plus first-order approximation as described in Section 3.3.

The RC, WD, and $\textsc{Sim-FOA}(\cdot)$ problems need to be solved iteratively. This can be treated as a fixed-point problem. A conventional way to solve a fixed-point problem is the method of successive averages (MSA).

In the typical system optimal **traffic** assignment problem, the optimal flow pattern is reached when for every OD pair, the marginal costs of all paths for this OD pair are the same. This implies that, ideally, when the flow distribution is optimal, we should have $\beta_{hkr}(\tilde{\boldsymbol{f}}) = \beta_{hkr'}(\tilde{\boldsymbol{f}})$ for all $r, r' \in R_k \setminus R_k^{\text{NoFlow}}$, where $R_k^{\text{NoFlow}} = \{r \in R_k \mid f_{hkr} = 0\}$ is the path set with zero flows. This implies that at the system optimal assignment, the marginal cost (travel time) of every non-zero flow path is the same (i.e., one cannot decrease the system travel time by switching passengers from one path to another).

However, in our study, this cannot be set as the convergence criterion because, in the dynamic **transit** assignment context, the cost function is not continuous due to left behind. Adding one more passenger to a path may lead to the system travel time increased by one or more headways. The following example illustrates that $\beta_{hkr}(\tilde{\boldsymbol{f}})$ can be arbitrarily large, which may cause the criterion of $\beta_{hkr}(\tilde{\boldsymbol{f}}) = \beta_{hkr'}(\tilde{\boldsymbol{f}})$ never being satisfied.

Example 1. *Consider a single direction bus line with N stations (Figure 3) and a fixed headway W. Assume every bus has a capacity of 1. There is one passenger waiting at each station except for the first station (i.e., there are $N-1$ waiting passengers). Now assume that one more passenger is added to station 1. Since the capacity of buses is 1, the newly added passenger will force all waiting passengers to be left behind one more time. Hence, the total added system travel time is $(N-1) \times W$. In this scenario, the $\beta_{hkr}(\tilde{\boldsymbol{f}})$ associated with the added passenger can be arbitrarily large depending on the number of stations N.*

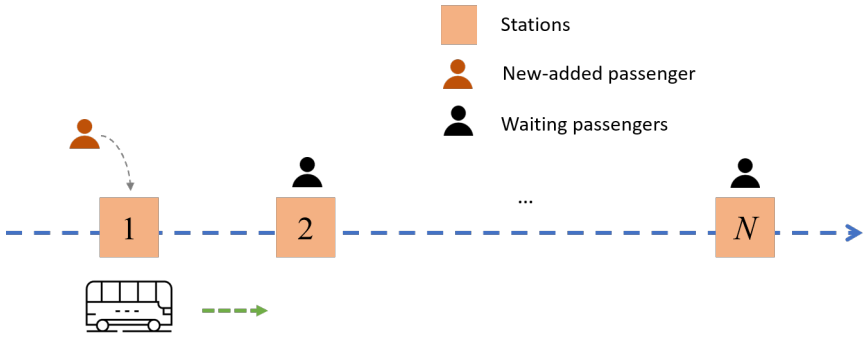


Figure 3: Example for arbitrarily large $\beta_{hkr}(\tilde{\boldsymbol{f}})$

Therefore, in this study, we define the convergence criteria based on the value of system travel time (i.e., when the value of the system travel time is relatively stable within a range). Specifically, it is assumed that the MSA algorithm has converged if

$$\left| Z(\tilde{\boldsymbol{f}})^{(n)} - \frac{1}{N^{\text{Cvg}}} \sum_{n'=n-N^{\text{Cvg}}}^{n-1} Z(\tilde{\boldsymbol{f}})^{(n')} \right| \leq \epsilon \tag{34}$$

where $Z(\tilde{\boldsymbol{f}})^{(n)}$ is the system travel time at the n-th iteration and ϵ is a predetermined threshold. Eq. 34 means that when the current system travel time is close to its average value of the last N^{Cvg} iterations, the algorithm terminates. Taking the average of the last N^{Cvg} iterations can mitigate the impact of fluctuations caused by the discontinuity of the system travel time.

The whole solution algorithm is described in Algorithm 1. Line 6 indicates the MSA step. Lines 10 and 11 mean that we will use the path shares with the smallest system travel time over the last $N^{\text{Cvg}} + 1$ iterations.

Algorithm 1 Solution procedure of the robust optimal flow problem

Initialize $\boldsymbol{p}^{(0)}$ (e.g., uniform path shares), $\boldsymbol{d}^{(0)}$ (e.g., nominal demand) and specify N^{Cvg}, ϵ.
Set iteration counter $n = 0$.
do
    $Z(\tilde{\boldsymbol{f}})^{(n)}, \boldsymbol{\beta}(\tilde{\boldsymbol{f}})^{(n)} = \textsc{Sim-FOA}(\boldsymbol{d}^{(n)}, \boldsymbol{p}^{(n)})$
    Solve the RC problem (Eq. 31) with $Z(\tilde{\boldsymbol{f}})^{(n)}$ and $\boldsymbol{\beta}(\tilde{\boldsymbol{f}})^{(n)}$ as inputs, and return $\hat{\boldsymbol{p}}^{(n+1)}$
    $\boldsymbol{p}^{(n+1)} = \frac{1}{n+1}\hat{\boldsymbol{p}}^{(n+1)} + (1 - \frac{1}{n+1})\boldsymbol{p}^{(n)}$
    Solve the WD problem (Eq. 32) with $\boldsymbol{p}^{(n+1)}$ as input and return $\boldsymbol{d}^{(n+1)}$
    $n = n + 1$
while $n \leq N^{\text{Cvg}}$ or $\left| Z(\tilde{\boldsymbol{f}})^{(n)} - \frac{1}{N^{\text{Cvg}}} \sum_{n'=n-N^{\text{Cvg}}}^{n-1} Z(\tilde{\boldsymbol{f}})^{(n')} \right| > \epsilon$
$n^* = \arg\min_{n'=n-N^{\text{Cvg}},\ldots,n} Z(\tilde{\boldsymbol{f}})^{(n')}$
return $\boldsymbol{p}^{(n^*)}$

Let $\boldsymbol{p}^*$ be the optimal path shares by from Algorithm 1. To realize the optimal path shares in the real world, the following system design can be used:

- Transit operators deploy the recommendation system to smartphone apps, websites, and electrical screens at stations.
- Passengers, when using the system, input their origins, destinations, and departure times.
- For a passenger input OD pair k and departure time h, the system will return a single recommended path r to them with probability p^*_{hkr}.

In this way, we expect the final path flows are close to the system optimal path flows if passengers follow the recommendation. In reality, passengers may have different preferences for different recommendations. That is, they may not follow the recommendations if they are provided with an unpreferred path. Appendix C discusses how to solve another path-passenger matching problem that incorporates passengers' preferences.

4. Model extensions

4.1. Solving the model in a rolling horizon

The model discussed in the previous section is a one-shot solution for path recommendation, which means the model will be run at the beginning of an incident (h_0) and output the recommendations for the whole period of interest $[h_0, h_H]$. In application, the model would be implemented in a rolling horizon framework.

Specifically, at time interval $\tilde{h}$, we first update the demand and supply information, including new demand estimates, new demand uncertainty sets, new available path sets, new service routes and frequencies, new incident duration estimates, etc. Based on the formulation above (i.e., let $h_0 = \tilde{h}$)), we solve the model to obtain recommendations for time $[\tilde{h}, h_H]$. But we only implement the recommendation strategies for the current time $\tilde{h}$ (i.e., $p^*_{\tilde{h}kr}$). An illustration of the rolling horizon implementation is shown in Figure 4. In this way, the new information obtained with the evolution of the incident and system operations can be used to improve model performance (this is known as adaptive RO).

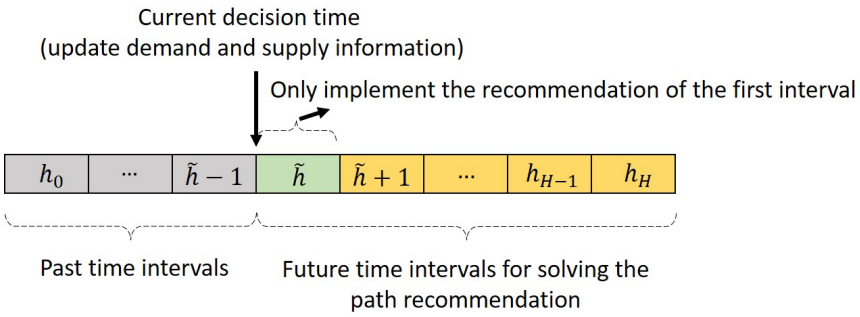


Figure 4: Illustration of the rolling horizon implementation

In the case study, we did not implement the rolling horizon for the following reasons. 1) First, as we use the CTA system as the case study, the operators in the selected incident did not update operation changes once the decisions have been made. Hence, it becomes less meaningful for the rolling horizon with fixed operations. 2) Second, there are computational time challenges in the case study. The main bottleneck comes from the simulation model. Running a simulation in the large-scale CTA network takes around 0.6 minutes. And our algorithm requires around 35 iterations to converge (see Section 6.1). 3) Third, as mentioned before, a more holistic implementation of the rolling horizon includes the update of the "uncertainty set", which implies an adaptive robust optimization. This requires additional derivations on how to use previous demand realizations to generate the new demand uncertainty set, which deserves separate future research.

4.2. Incident duration uncertainty

In this study, we assume operators have a reasonable estimate of incident duration. However, it is possible that we can only obtain a distribution of incident duration. In this section, we show that our formulation can be easily extended to capture the incident duration uncertainty with stochastic optimization (SO) techniques[4].

Let the set of all possible incident scenarios be Ω. For example, Ω may include incidents with a duration of 30, 40, or 50 minutes. For each scenario $\xi \in \Omega$, we denote $\beta_{hkr}(\tilde{\boldsymbol{f}};\xi)$ and $Z(\tilde{\boldsymbol{f}};\xi)$ as the approximated gradient and current system travel time under flow $\tilde{\boldsymbol{f}}$ and incident scenario ξ. Hence, the objective function for the RO problem becomes:

$$\begin{aligned}\mathbb{E}[\hat{Z}(\boldsymbol{p},\boldsymbol{u},\boldsymbol{v},\boldsymbol{y})^{\mathrm{RC}}] = \sum_{\xi\in\Omega}\mathbb{P}(\xi)\left[Z(\tilde{\boldsymbol{f}};\xi)+\sum_{(h,k,r)\in\mathcal{F}}\beta_{hkr}(\tilde{\boldsymbol{f}};\xi)\cdot(d_{hk}\cdot p_{hkr}-\tilde{f}_{hkr})\right]+\rho_{1-\varepsilon}\left\|\boldsymbol{y}_1\right\|_2 \\ +(\boldsymbol{d}^{\mathrm{U}}-\bar{\boldsymbol{d}})^T\boldsymbol{u}_1+(\bar{\boldsymbol{d}}-\boldsymbol{d}^{\mathrm{L}})^T\boldsymbol{u}_2+(\boldsymbol{d}^{\mathrm{U}}_{\mathcal{H}}-\boldsymbol{S}\bar{\boldsymbol{d}})^T\boldsymbol{v}_1+(\boldsymbol{S}\bar{\boldsymbol{d}}-\boldsymbol{d}^{\mathrm{L}}_{\mathcal{H}})^T\boldsymbol{v}_2+(\Gamma-1)\cdot(\mathbf{1}^T\bar{\boldsymbol{d}})\cdot u_3\end{aligned} \tag{35}$$

where $\mathbb{P}(\xi)$ is the probability of scenario ξ being realized. The expectation above is taking over different incident scenarios. Define $Z(\tilde{\boldsymbol{f}};\Omega) := \sum_{\xi\in\Omega}\mathbb{P}(\xi)Z(\tilde{\boldsymbol{f}};\xi)$ and $\beta_{hkr}(\tilde{\boldsymbol{f}};\Omega) := \sum_{\xi\in\Omega}\mathbb{P}(\xi)\beta_{hkr}(\tilde{\boldsymbol{f}};\xi)$,

[4]The reason for using SO, instead of RO, to capture incident duration uncertainty is that the worst-case scenario for the incident duration is always the largest one, which makes the problem trivial and may not reflect reality.

substituting them into the objective function

$$\mathbb{E}[\hat{Z}(\boldsymbol{p},\boldsymbol{u},\boldsymbol{v},\boldsymbol{y})^{\mathrm{RC}}] = \sum_{(h,k,r)\in\mathcal{F}} \beta_{hkr}(\tilde{\boldsymbol{f}};\Omega)\cdot(d_{hk}\cdot p_{hkr} - \tilde{f}_{hkr}) + \rho_{1-\varepsilon}\left\|\boldsymbol{y}_1\right\|_2 + (\boldsymbol{d}^{\mathrm{U}} - \bar{\boldsymbol{d}})^T\boldsymbol{u}_1$$
$$+(\bar{\boldsymbol{d}} - \boldsymbol{d}^{\mathrm{L}})^T\boldsymbol{u}_2 + (\boldsymbol{d}^{\mathrm{U}}_{\mathcal{H}} - \boldsymbol{S}\bar{\boldsymbol{d}})^T\boldsymbol{v}_1 + (\boldsymbol{S}\bar{\boldsymbol{d}} - \boldsymbol{d}^{\mathrm{L}}_{\mathcal{H}})^T\boldsymbol{v}_2 + (\Gamma-1)\cdot(\boldsymbol{1}^T\bar{\boldsymbol{d}})\cdot u_3 + Z(\tilde{\boldsymbol{f}};\Omega) \quad (36)$$

As the constraints in the RO problem are not related to incident scenarios (i.e., $\beta_{hkr}(\tilde{\boldsymbol{f}})$ and $Z(\tilde{\boldsymbol{f}})$ are not included in the constraint part), this implies that incorporating the incident duration uncertainty with SO only requires a change in the objective function.

5. Case study design

In the case study, we consider an actual incident in the Blue line of the Chicago Transit Authority (CTA) urban rail system (Figure 5). The incident starts at 8:14 AM and ends at 9:13 AM on Feb 1st, 2019 due to infrastructure issues between Harlem and Jefferson Park stations. The entire Blue Line was suspended. During the disruption, the Loop (Chicago CBD area) is the destination for most passengers. Usually, there are four paths leading to the Loop: 1) using Blue Line (i.e., waiting for the system to recover), 2) using the parallel bus lines, 3) using the North-South (NS) bus lines to transfer to the Green Line, and 4) using the West-East (WE) bus lines to transfer to the Brown Line. Based on the service structure, we can construct the route sets R_k for each OD pair k.

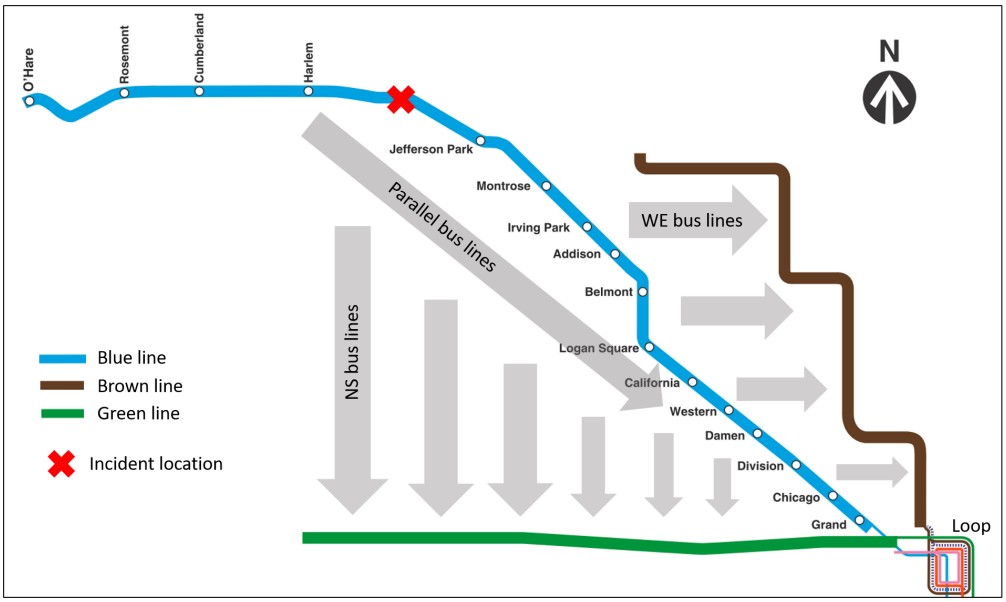


Figure 5: Case study diagram

5.1. Parameter setting

$\mathcal{K}$ is the set of all OD pairs with origins at the Blue Line and destinations at the Loop. The response time is set as $\eta = 0$ for simplicity (i.e., assuming a quick response for operators). The time interval is set to $\tau = 10$ mins. The time period with recommendation is set as $h_H = 10$, corresponding to 9:44 - 9:54 AM (i.e., 50 minutes after the end of the incident). In this study, we assume that the incident duration is known

or can be reasonably estimated. The factor of total demand level Γ is set to 1.1, which is the 90% percentile of the total demand distribution.

The validation of the simulation model's performance is shown in Appendix D. Results show that the model can capture the passenger and vehicle interactions well in the CTA system.

5.2. Quantification of uncertainty sets

The demand uncertainty is determined by the nominal demand $\bar{\boldsymbol{d}}$, covariance matrix $\boldsymbol{\Sigma}$ (which can be used to get $\boldsymbol{D}$), and upper and lower bounds for demand (i.e., $\boldsymbol{d}^{\mathrm{U}}$, $\boldsymbol{d}^{\mathrm{L}}$, $\boldsymbol{d}^{\mathrm{U}}_{\mathcal{H}}$, $\boldsymbol{d}^{\mathrm{L}}_{\mathcal{H}}$). These can be estimated from historical demand. However, as the demand on the incident day is smaller than usual given that some passengers may leave the system, we cannot directly use normal day smart card data as historical demand. One possible solution is to use data from previous days with similar incidents. Nevertheless, this is usually unavailable due to the lack of enough similar incidents. Hence, in this study, we first use survey results and historical smart card data to generate "synthetic historical demand" samples, and then estimate the uncertainty set from the samples.

There are two sources of demand uncertainty: 1) the inherent demand variations across different days and 2) the uncertainty of how many passengers left the PT system during the incident. The first part can be captured by historical smart card data (without incidents). The second part is approximated by the survey results. According to previous survey-based studies, the proportion of the passengers leaving the PT system during incidents is around 10%∼30% (Lin et al., 2016; Rahimi et al., 2020). Then, the "synthetic historical demand" is generated as follows:

- Collect smart card data from a recent workday and calculate the demand vector without passengers leaving the system for each (h, k) (the demand for $h = 0$, i.e., offloading passengers, can be obtained using the simulation model).
- For each (h, k), we randomly draw a proportion of leaving passengers from a uniform distribution $\mathcal{U}(10\%, 30\%)$[5]. The demand after removing the leaving passengers is the incident period demand vector.

We collected a total of 16 weekdays from Jan 2019 (the previous month of the incident day) and generated 16 sample demand vectors. The mean value is used as the nominal demand $\bar{\boldsymbol{d}}$ and the covariance matrix $\boldsymbol{\Sigma}$ is estimated from these samples. The upper and lower bounds for demand (i.e., $\boldsymbol{d}^{\mathrm{U}}$, $\boldsymbol{d}^{\mathrm{L}}$, $\boldsymbol{d}^{\mathrm{U}}_{\mathcal{H}}$, $\boldsymbol{d}^{\mathrm{L}}_{\mathcal{H}}$) are set as the samples' maximum and minimum values, respectively.

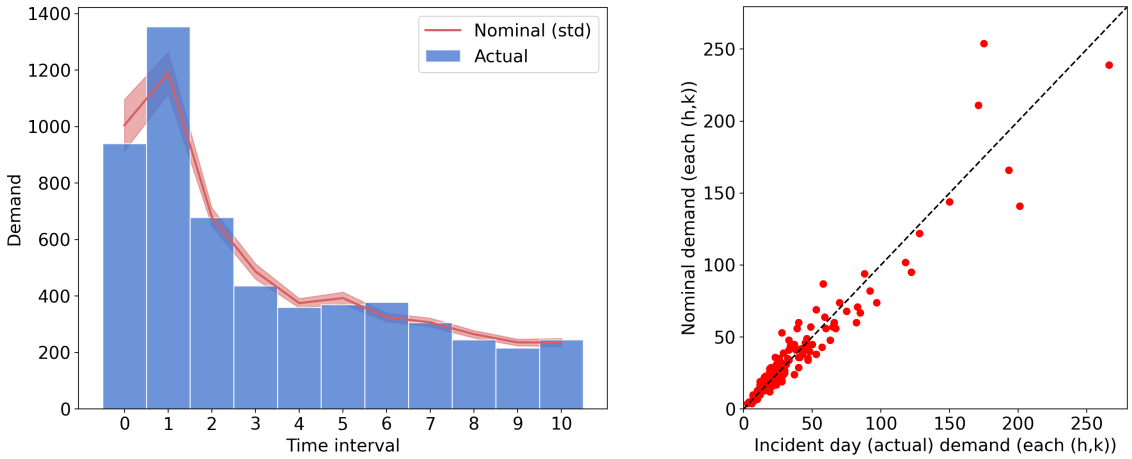
The hyperparameter $\rho_{1-\varepsilon}$ for the ellipsoidal uncertainty set are chosen from these values: $\{0, 0.25, 0.52, 0.84, 1.28, 1.64, 2.33\}$, which corresponds to the $\{50, 60, 70, 80, 90, 95, 99\}$ percentiles of the standard normal distribution. Note that $\rho_{1-\varepsilon} = 0$ represents the case of no uncertainty (i.e., nominal model).

5.3. Data description

The nominal and actual (incident day) demand comparison is shown in Figure 6. The total nominal demand is 5,499, similar to the total actual demand (5,531), implying that introducing a proportion of leaving passengers (i.e., 10% - 30%) can capture the demand reduction on the incident day. We also observe that the aggregate nominal demand for each time interval is similar to that of the incident day. The major differences

[5]We use uniform distribution because we have no distributional information of the leaving passenger proportions

happen at the first two time intervals ($h = 0, 1$). However, looking at the demand for each (h, k) (Figure 6b), the differences are more prominent. The discrepancy between nominal and actual demands indicates the potential for the RO approach to perform better.



(a) Total demand for each time interval h

(b) Demand comparison for each (h, k)

Figure 6: Demand patterns

Table 1 shows the results of the Mardia test of multivariate normality (Cain et al., 2017) for demand samples. The Mardia test is used to check whether the sample's multivariate skewness and kurtosis are consistent with a multivariate normal distribution. If both are satisfied, we can assume the samples are multivariate normally distributed. We observe that, in Table 1, the synthetic historical demands have consistent skewness but inconsistent kurtosis with the multivariate normal distribution, suggesting that they are not multivariate normally distributed. However, as skewness is a measure of the asymmetry of the probability distribution of a random variable about its mean, the Mardia Skewness testing shows that the demand distribution is symmetric. Hence, it is still reasonable to use the ellipsoidal uncertainty set to describe a symmetric distributed random variable. Moreover, as mentioned in Remark 1, the distribution of a variable does not affect the definition of the uncertainty set (it only affects the calculation of probability guarantees).

Table 1: Mardia test of multivariate normality

Test	p-value	Test	p-value
Mardia Skewness	1.00	Mardia Kurtosis	0.00

Note: The null hypothesis is that the samples are multivariate normally distributed. A small p-value indicates we are more likely to reject the null hypothesis.

5.4. Benchmark models

The following approaches are used to obtain benchmark path shares.

Uniform path shares. The uniform path shares are defined as $p_{hkr} = \frac{1}{|R_k|} \; \forall \; r \in R_k$. This is a naive model corresponding to the intuition of "distributing passengers to different paths" when no information is

available.

Capacity-based path shares. The capacity-based path shares aim to assign passengers to different paths according to the path capacity. Specifically, for a path r in OD pair k and time h, we calculate the path capacity as the total available capacity of all vehicles passing through the first boarding station of the path (denoted as C_{hkr}). The capacity-based path shares are defined as

$$p_{hkr} = \frac{C_{hkr}}{\sum_{r \in R_k} C_{hkr}} \quad \forall\, r \in R_k, h \in \mathcal{H}, k \in \mathcal{K}, \tag{37}$$

For example, for a path consisting of an NS bus route and the Green Line, C_{hkr} is calculated as the total available capacity of all buses at the boarding station of the NS bus route during time interval h. The available capacity can be obtained from the simulation model using historical demand. The available capacity for the Blue Line (i.e., incident line) depends on the revised schedules during the incident (i.e., the service suspension is considered). When no trains operate on the Blue Line, the corresponding C_{hkr} will be zero.

Status-quo path shares. The status-quo path shares are the inferred path choices of passengers on the incident day. During the incident period, the demand on the WE, NS, and parallel bus lines experienced an increase. The difference from the average demand on normal days can be seen as the number of passengers choosing the corresponding path. Hence, by identifying the demand increase for all nearby bus stops, we can get the number of passengers using the parallel bus, NS+Green, and WE+Brown paths for each OD pair k and time interval h. However, the number of waiting passengers in the Blue Line cannot be directly inferred because the CTA system does not record the tap-out information. Hence, we approximate the proportion of waiting passengers based on survey results (Rahimi et al., 2019). Rahimi et al. (2019) used a survival model to analyze the waiting time tolerance of CTA riders during a service disruption. The model results provide the proportion of waiting passengers given different system recovery times. Therefore, the status-quo path shares are inferred as follows:

- Step 1: Given the current time interval h and the incident end time T_e, the remaining time until the end of the incident is $T_e - h$. Therefore, if passengers choose to wait, their waiting time will also be $T_e - h$. Based on the hazard model in Rahimi et al. (2019), we can obtain the proportion of waiting passengers given the waiting time, denoted as $p_{\text{wait}}(T_e - h)$.
- Step 2: For each OD pair k and time interval h, the number of passengers using the parallel bus, NS+Green, and WE+Brown paths can be calculated based on demand increase compared to the normal demand. Let the demand increase for path r of OD pair k at time h be DI_{hkr}, where $r \in R_k \setminus \{r_{\text{wait}}\}$, r_{wait} represents the path of waiting for the Blue Line.
- Step 3: The status quo path shares are calculated as follows:

$$p_{hkr_{\text{wait}}} = p_{\text{wait}}(T_e - h) \quad \forall\, h \in \mathcal{H}, k \in \mathcal{K}, \tag{38}$$

$$p_{hkr} = (1 - p_{hkr_{\text{wait}}}) \cdot \frac{DI_{hkr}}{\sum_{r \in R_k \setminus \{r_{\text{wait}}\}} DI_{hkr}} \quad \forall\, r \in R_k \setminus \{r_{\text{wait}}\},\ h \in \mathcal{H}, k \in \mathcal{K} \tag{39}$$

6. Results

In this section, we demonstrate the model's performance in two steps. In the first step, results of the optimization model without uncertainty (i.e., the nominal model with $\rho_{1-\epsilon} = 0$) are compared with the

three benchmark path shares. In the second step, we compare the results from the robust model with the results from the nominal model in order to assess the value of considering uncertainties in generating path recommendations.

6.1. Model convergence and computational time

Figure 7 shows the convergence of the nominal ($\rho_{1-\epsilon} = 0$) and robust (with $\rho_{1-\epsilon} = 0.84$) models. The simulation-based linearization and MSA successfully decrease the system travel time. The model converges within 35 iterations. Note that the optimal cost for the robust model is higher than the nominal model. This is expected since the robust model assumes the worst-case demand (by definition with higher system travel time). The performance of the corresponding path recommendations will be evaluated based on the actual demand (discussed in the next section).

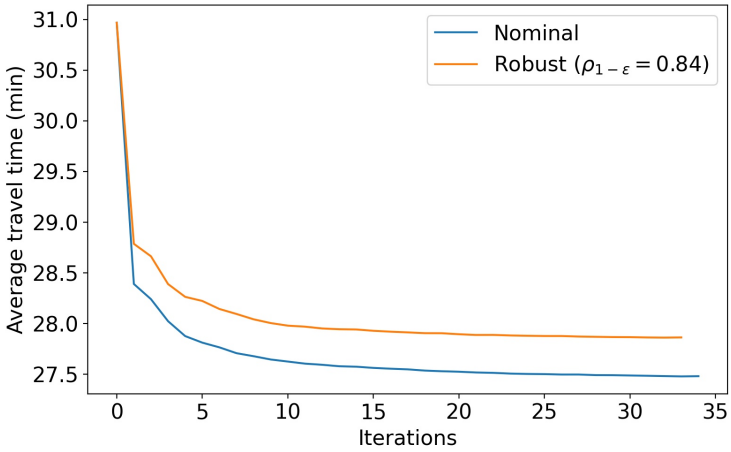


Figure 7: Convergence of optimization models

The mode's computational time mainly depends on the speed of the simulation and the number of iterations. Solving Eqs. 31 and 32 is quite efficient because of the tractable SOCP formulations. Currently, running one simulation for the CTA system for 2 hours takes around 0.6 minutes. Hence, the total solving time for a RO model (assuming 35 iterations) is around 23 minutes.

The computational time of 23 minutes is too long for real-world rolling horizon implementation. However, since the bottleneck is the simulation process, future studies can improve the model's efficiency by enhancing the speed of the simulation model (e.g., code with C++). If the simulation time is reduced to 5 seconds, the total solution time will be reduced to around 3 minutes, which enables the real-world rolling horizon implementation with 15 minutes intervals (including times for updating inputs). Another way to simplify the model is to reduce the look ahead horizons. Now we consider a 2-hour horizon until the end of the incident. With rolling horizon implementation, since the incident information will be updated in real-time, we may only need to look ahead for 1 hour and wait for new information to come. Reducing the look-ahead time can significantly simplify the model and reduce the computational time.

6.2. Model evaluation

The optimization model only utilizes information about the nominal demand and the associated uncertainty set. The actual demand is unknown when running the model (otherwise there are no uncertainties). After obtaining the path shares (either from optimization or the benchmark models), the recommendation

strategies are evaluated based on the actual incident day demand using the simulation model. We assume passengers would follow the path recommendation. The simulation model can output the travel times of every passenger in the system, and can be used to compare the performance of the path shares obtained from the various approaches. Performance is measured in terms of average travel time and average waiting time.

6.3. Nominal vs. Benchmark models

Table 2 compares the results for different path shares, The result of the no incident scenario is also shown for comparison. The average travel times are calculated over all passengers (a total of 27,007 passengers) and the passengers who originally planned to use the Blue Line (i.e., passengers who are provided with recommendations, a total of 5,531 passengers, a subset of the 27,007 passengers). Results show that the optimization-based path shares outperform all benchmark models. For all passengers in the system, the average travel time is reduced by 9.1% compared to the status quo. And for the incident line passengers, the reduction is even higher (20.6%).

Recommendations based on the uniform path shares result in worse performance than the status quo scenario. This implies that current passengers' choices are not random and show some rationality. The capacity-based path shares can also reduce the system travel time significantly (by 6.9%). However, as the capacity-based path recommendations do not capture the spatial and temporal changes in available capacity due to passenger flow redistribution, they are worse than the optimization-based results. A more comprehensive discussion on the performance comparison between the optimization model and the capacity-based model can be found in Appendix E.

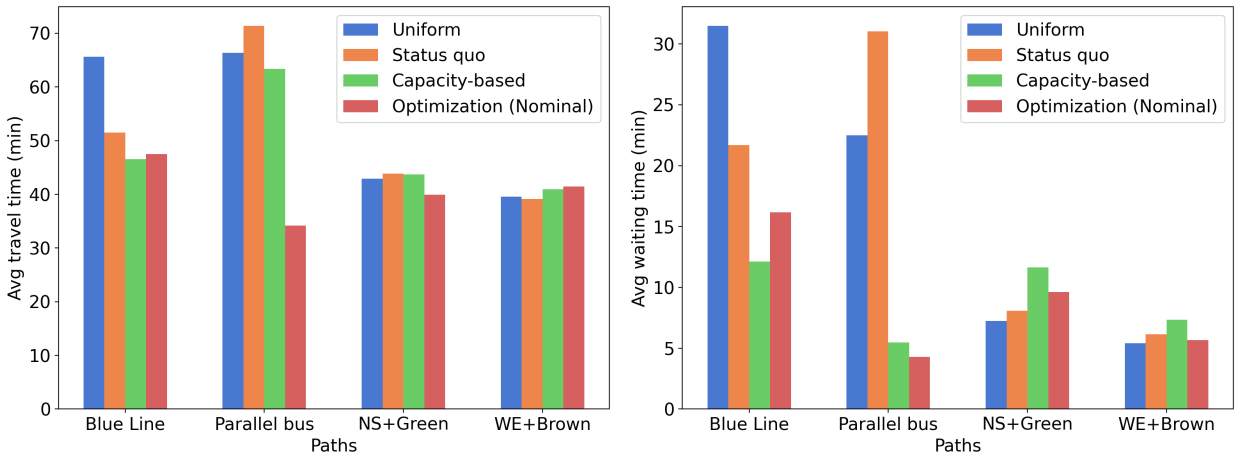
Compared to the no-incident scenario, we find that the influence of incidents is significant. Path recommendations can only alleviate the impact of service disruption but are far from eliminating. Even with the optimization-based path recommendations, we still have more than two times of travel time for incident-line passengers compared to the no-incident situation.

Table 2: Average travel time comparison

Scenarios	All passengers (# 27,003)		Incident-line passengers (# 5,531)	
	Avg travel time (min)	% change[1]	Avg travel time (min)	% change[1]
No incident	21.81	-	18.95	-
Uniform	31.02	+1.7%	54.64	+6.4%
Status quo	30.49	0%	51.34	0%
Capacity-based	28.36	-6.9%	43.23	-15.8%
Optimization (nominal)	27.71	**-9.1%**	40.75	**-20.6%**

[1]: changes compared to the status quo scenario

Figure 8 shows the average travel time and waiting time for different paths for all incident line passengers. We observe that the optimization-based path recommendations have more consistent travel time across the four types of paths, implying a better utilization of the system's capacity. However, for other recommendation strategies, passengers using parallel buses have significantly longer travel times than those using other alternatives. Figure 8 also shows that the average waiting time for the status quo scenario is around 30 minutes, which means most passengers chose to use the parallel bus during the incident, causing severe congestion. However, with the optimization-based path shares, the average waiting time for the parallel bus is less than 5 minutes (around a headway).



(a) Average travel time for different paths

(b) Average waiting time for different paths

Figure 8: Comparison of average travel time and waiting time of different paths for incident line passengers

The objective of this study is to minimize the system travel time. However, under the optimal path shares, some passengers' travel time may be increased compared to the status quo. Figure 9 shows the distribution of changes in individual travel time (optimization-based minus the status quo) for all passengers whose path choice under the recommendation scenario is different than their choice in the status quo scenario. Most passengers experience lower travel times. However, some passengers become worse off after following the path recommendations. This is a typical drawback of system optimal (first-best) assignment (Lawphongpanich and Yin, 2010). Future studies may explore a Pareto-improving (second-best) path recommendation that ensures no individual becomes worse-off. In reality, when implementing the recommendations, some paths that lead to extremely worse travel time compared to the status quo can be dropped from the solution.

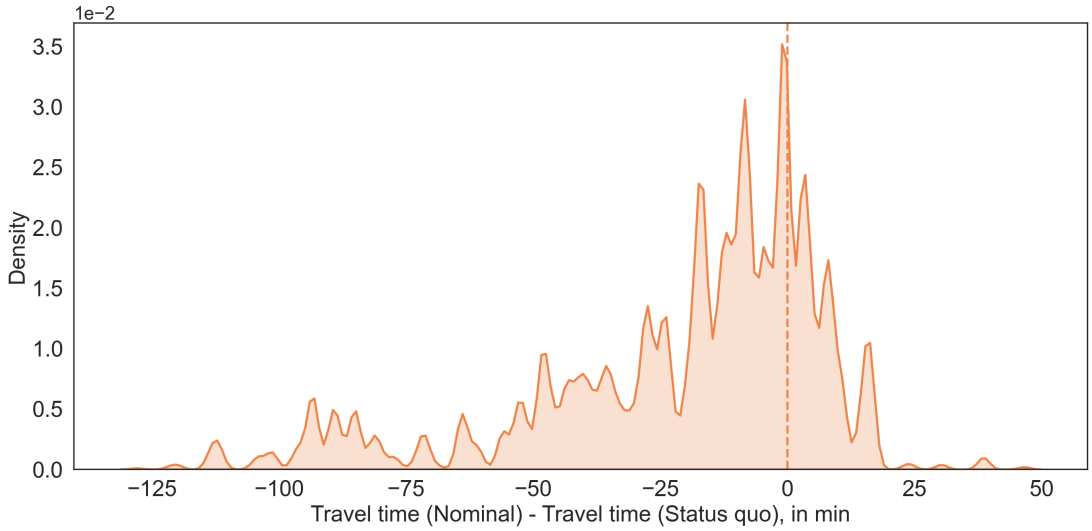


Figure 9: Distribution of the change in individual travel time (not including passengers without changes as they will distort the distribution with too much density concentrated at zero)

6.4. Robust models vs. Nominal model

6.4.1. Model comparison under actual demand

Figure 10 compares the results, in terms of travel time, of the RO approach with different values of $\rho_{1-\epsilon}$ under the actual demand. For all values of the robust model except for $\rho_{1-\epsilon} = 2.33$, the RO approach shows better performance than the nominal model. This implies that considering the demand uncertainty in determining the recommendation can further improve the effectiveness of path recommendation strategies. The best value is $\rho_{1-\epsilon} = 0.84$, where the travel time for the incident line passengers is reduced by 2.91% compared to the nominal model. Note that the percentage decreases are relatively small because some passengers' travel times are not changed. If we only look at incident-line passengers with travel time changes, the average travel times are 47.6 min and 37.9 min for the nominal and RO ($\rho_{1-\epsilon} = 0.84$) scenarios, respectively, where the travel time reductions are 20.4%.

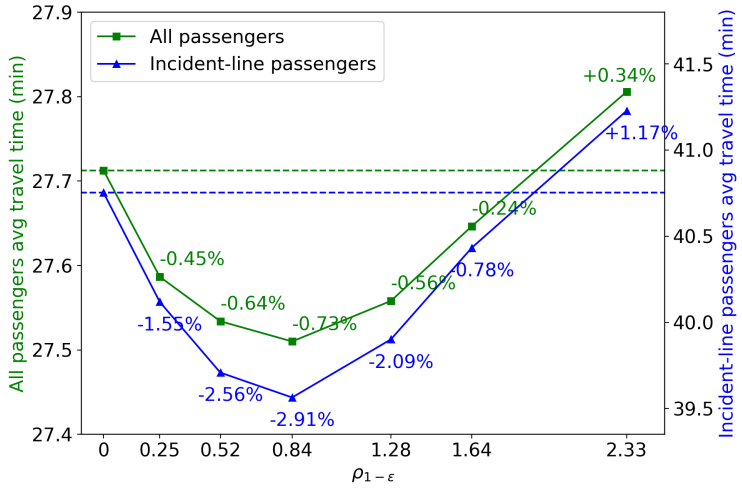


Figure 10: Performance of RO. The percentage changes are compared to the nominal scenario

Note that using $\rho_{1-\epsilon} = 2.33$ results in the largest uncertainty set compared to other values. This reflects a very conservative scenario where the agency prefers to plan against a very high realization of demand. In this case, the worst-case demand patterns may deviate from the actual demand too much, thus performing worse than the nominal model. Figure 11 illustrates the worst-case demand for different values of $\rho_{1-\epsilon}$. The worst-case demands for the $\rho_{1-\epsilon} = 0.52, 0.84, 1.28$ scenarios are closer to the actual demand, while $\rho_{1-\epsilon} = 2.33$ overestimates the demands, especially for the earliest periods ($h = 0, 1$) (which are the most critical periods). These results are consistent with the travel time performance in Figure 10.

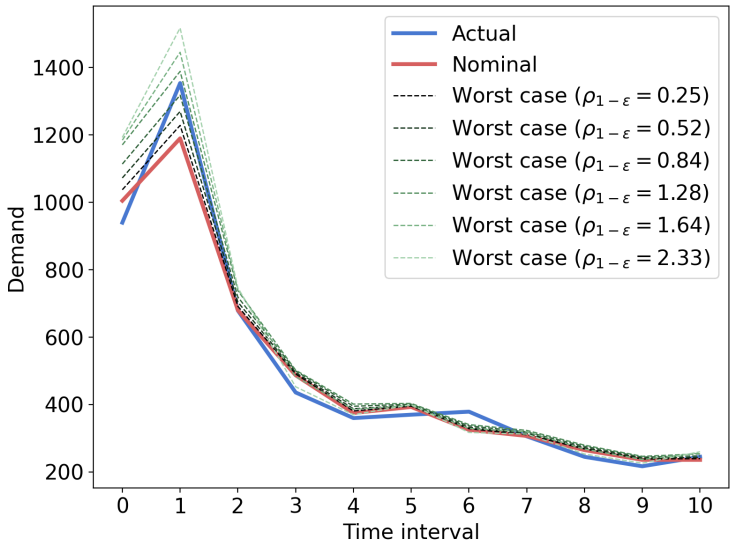


Figure 11: Worst-case demand patterns

6.4.2. Model comparison under random demand

To further validate the model's performance, we test the performance of the solution obtained from the RO approach on the 16 demand samples generated in Section 5.2. These demand samples represent different possible realizations of the incident day demand. Figure 12 shows the compassion of the random demand samples versus the actual and nominal demands. Notice that the random demand samples include both high and low demand scenarios, which can better validate the performance of the RO approach under different demand patterns.

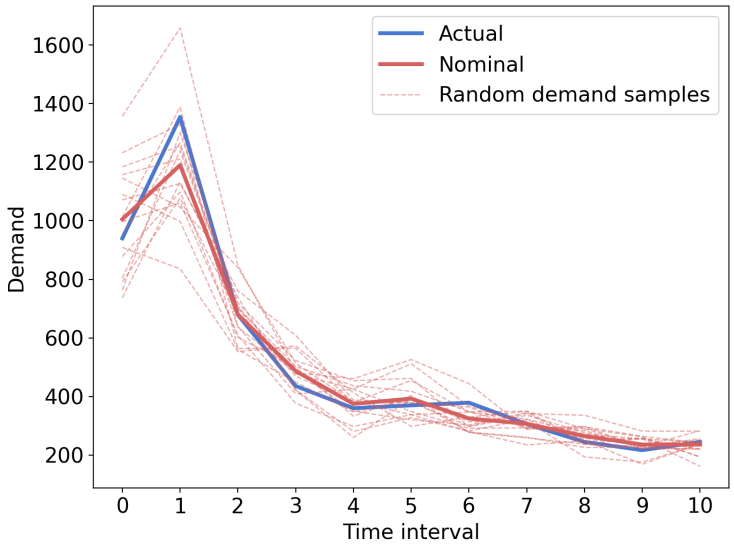


Figure 12: Random demand patterns for experiments

Table 3 compares the results of average travel time for different RO models. The numbers in the table are the mean values of the 16 experiments. The performances are similar to the results under the actual demand. The RO approach shows better performance than the nominal model for all values of $\rho_{1-\epsilon}$ the robust model

except for $\rho_{1-\epsilon} = 2.33$. The reasons may be that the RO approach focuses more on critical OD pairs and time intervals where the path recommendations for them are considered more important for system performance.

Table 3: Average travel time comparison for RO models

Models	All passengers		Incident-line passengers	
	Avg travel time (min)	% change[1]	Avg travel time (min)	% change[1]
Nominal ($\rho_{1-\epsilon} = 0$)	27.79	-	41.08	-
$\rho_{1-\epsilon} = 0.25$	27.70	-0.32%	40.57	-1.23%
$\rho_{1-\epsilon} = 0.52$	27.65	-0.48%	40.24	-2.05%
$\rho_{1-\epsilon} = 0.84$	27.64	-0.54%	40.13	-2.31%
$\rho_{1-\epsilon} = 1.28$	27.68	-0.39%	40.41	-1.62%
$\rho_{1-\epsilon} = 1.64$	27.74	-0.17%	40.83	-0.60%
$\rho_{1-\epsilon} = 2.33$	27.86	+0.27%	41.47	+0.96%

[1]: changes compared to the nominal model

6.4.3. Model comparison under worst-case demand

Theoretically, the RO model should show the best performance under the worst-case demand. Table X shows the comparison between the nominal model and RO models under the worst-case demand (the worst-case demand patterns are shown in Figure 11). For each value of $\rho_{1-\epsilon}$, we obtain the results for both the nominal model and the robust model by assuming the actual demand is the worst-case demand. Note that with a higher value of $\rho_{1-\epsilon}$, the worst-case demand patterns become worse and the average travel time of passengers will increase regardless of which model we use.

As shown in Table 4, if the "actual demand" (i.e., the demand for model evaluation) is the worst-case demand, the robust models consistently outperform the nominal models. And the improvement is higher than that of the previous experiments. Moreover, the higher the value of $\rho_{1-\epsilon}$ (i.e., more extreme demand patterns), the higher the improvement of robust models compared to the nominal models. This emphasizes the importance of considering demand uncertainties under extreme demand patterns.

Table 4: Average travel time comparison under the worst-case demands

Uncertainty set	Total demand	All passengers avg travel time (min)		Incident-line passengers avg travel time (min)	
		Nominal model	Robust model (% change[1])	Nominal model	Robust model (% change)
$\rho_{1-\epsilon} = 0.25$	27,084	27.99	27.67 (-1.14%)	41.77	40.43 (-3.21%)
$\rho_{1-\epsilon} = 0.52$	27,195	28.12	27.75 (-1.32%)	42.14	40.62 (-3.61%)
$\rho_{1-\epsilon} = 0.84$	27,344	28.26	27.86 (-1.41%)	42.47	40.89 (-3.72%)
$\rho_{1-\epsilon} = 1.28$	27,525	28.37	27.91 (-1.62%)	42.99	41.22 (-4.12%)
$\rho_{1-\epsilon} = 1.64$	27,522	28.61	28.05 (-1.96%)	43.81	41.74 (-4.72%)
$\rho_{1-\epsilon} = 2.33$	27,520	28.93	28.28 (-2.25%)	45.28	42.65 (-5.80%)

[1]: changes compared to the nominal model with the same worst-case demand

7. Conclusion and discussion

In this paper, we propose a path recommendation model to mitigate congestion during public transit disruptions. Passengers with different ODs and departure times are recommended alternative paths to use

such that the total system travel time is minimized. To tackle the non-analytical formulation of travel times due to left behind, we propose a simulation-based first-order approximation to transform the original problem into a linear program and solve the new problem iteratively with MSA. Uncertainties in demand are modeled using RO techniques to protect the path recommendation strategies against inaccurate estimates. A real-world rail disruption scenario in the CTA system is used as a case study. Results show that even without considering uncertainty, the nominal model can reduce the system travel time by 9.1% (compared to the status quo), and outperforms the benchmark capacity-based path recommendation. The average travel time of passengers in the incident line is reduced more (-20.6% compared to the status quo). After incorporating the demand uncertainty, the robust model further reduces the system travel time. The best robust model with $\rho_{1-\epsilon} = 0.84$ decreases the average travel time of incident-line passengers by 2.91% compared to the nominal model.

The performance improvement by incorporating demand uncertainty is not very significant. The reason may be that demand variations at the incident situation have a limited impact on the optimal path shares. Notice that the demand during an incident is already very high for the system (due to the reduced supply level). Hence, the path recommendation patterns under nominal and worst-case demand may be similar. However, the methodology presented in this study provides a general way to deal with PT demand uncertainty. It can be used for other operations control, optimization, planning, or recommendation applications.

Though we discussed potential model extensions with rolling horizon and incident duration uncertainty, we did not implement these extensions in the case study as the focus has been on the methodology for solving the problem. Incorporating real-time information as an adaptive RO would generally increase model performance (Bertsimas et al., 2011). This presents an interesting future research direction. Other future research directions include the following. 1) Current demand uncertainty sets need to be quantified with a budget factor $\rho_{1-\varepsilon}$. The choice of budget factor usually relies on numerical testing (Bertsimas et al., 2012; Guo et al., 2021). Future studies may also develop data-driven uncertainty quantification methods to automate the hyperparameter tuning task. 2) As shown in Figure 9, the system optimal path recommendation may result in worse-off travel time for some passengers, causing equity and fairness issues. Future studies may consider incorporating Pareto-improving constraints to ensure that all passengers are better-off if following our recommendation. 3) In this study, we assume that passengers follow the recommendation. Non-compliance, however, if present, may lead to the actual path flows deviating from the optimal ones. Future research may focus on approaches for path recommendations that capture behavior uncertainty. 4) Finally, this study presents an OD-based (aggregated) path recommendation regime. Passengers with the same OD and departure time are treated homogeneously. In reality, different passengers may have different preferences on path choices. And these preferences can affect their compliance with recommendations. Future studies can develop an individualized path recommendation system considering heterogeneous passenger preferences.

8. Authors' contribution

Baichuan Mo: Conceptualization, Methodology, Software, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Haris N. Koutsopoulos**: Conceptualization, Supervision, Formal analysis, Writing - Review & Editing, **Zuo-Jun Max Shen**: Conceptualization, Supervision. **Jinhua Zhao:** Conceptualization, Supervision, Project administration, Funding acquisition.

9. Acknowledgement

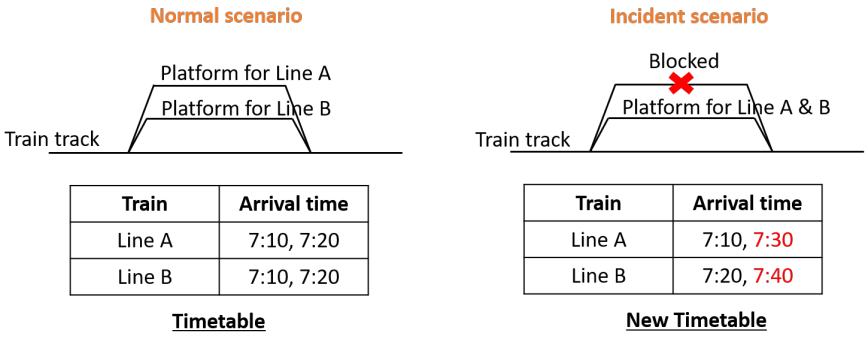
The authors would like to thank Chicago Transit Authority (CTA) for their support and data availability for this research.

Appendices

Appendix A. Capturing supply changes by adjusting timetable

Since all the operations can be described by timetable, it is reasonable to capture the supply changes in disruptions by adjusting the timetable. In this appendix, we show that our simulation model is able to capture the "partially blocked tracks" at a platform with complex configurations (e.g., different train capacities or different lines). That is, even if only one specific track failed in a platform, this type of disruption can be captured by the change in vehicles' timetables.

For example, consider a platform with 2 different lines A and B (Figure A.13). Suppose that only the track associated with Line A is blocked. And operators decide to let Lines A and B share the same remaining track. This operation change can be captured by adjusting the train's timetable for Lines A and B (i.e., trains may have higher headway, and they cannot use the platform simultaneously in the new timetable). In this example, the second vehicle of Line A is delayed to 7:30 and the second vehicle of Line B is delayed to 7:40.



Train	Arrival time
Line A	7:10, 7:20
Line B	7:10, 7:20

Timetable

Train	Arrival time
Line A	7:10, 7:30
Line B	7:20, 7:40

New Timetable

Figure A.13: Example of representing supply changes as timetable changes in a multi-platform scenario

Appendix B. Simulation-based first-order approximation

Appendix B.1. Calculation of $T^{Q}_{hkr}(\tilde{\boldsymbol{f}})$

Let $\mathcal{V}^b_{hkr}$ be the set of vehicles that the $\mathcal{M}_{hkr}$ passengers board at station b. Adding an additional passenger to $\mathcal{M}_{hkr}$ means one more passenger boards one of the vehicles in $\mathcal{V}^b_{hkr}$. Let $\mathbb{1}_{\{\text{Full}^b_v\}}$ be an indicator of whether vehicle v is full or not after its departure from station b. Then the total increase in system travel time for passengers queuing behind $\mathcal{M}_{hkr}$ is:

$$T^{\mathrm{Q}}_{hkr}(\tilde{\boldsymbol{f}}) = \sum_{b\in\mathcal{B}_{hkr}} \sum_{v\in\mathcal{V}^b_{hkr}} \frac{\mathbb{1}_{\{\text{Full}^b_v\}} \cdot W^b_v}{|\mathcal{V}^b_{hkr}|} \tag{B.1}$$

where $\mathcal{B}_{hkr}$ is the set of all boarding stations for $\mathcal{M}_{hkr}$ passengers (in the example of Figure 2, $\mathcal{B}_{hkr} = \{a_1, a_5\}$). W_v^b is the headway of vehicle v at station b. The sum over all vehicles is because we do not specify the exact vehicle that the additional passenger will board, and thus take the average over all vehicles. In the example of Figure 2, since there are two boarding stations for $\mathcal{M}_{hkr}$ (a_1, a_5), $T_{hkr}^{\mathrm{Q}}(\tilde{\boldsymbol{f}})$ is approximately two headways if the vehicles are full.

Appendix B.2. Calculation of $T_{hkr}^{O}(\tilde{\boldsymbol{f}})$

Let $\mathcal{O}_{hkr}^v$ be the set of all on-board stations for $\mathcal{M}_{hkr}$ and vehicle $v \in \mathcal{V}_{hkr}^b$. For example, for vehicles in Line 1 in Figure 2, $\mathcal{O}_{hkr}^v$ will be a_2, a_3, and a_4. Then the travel time increase for passengers waiting at on-board stations is:

$$T_{hkr}^{\mathrm{O}}(\tilde{\boldsymbol{f}}) = \sum_{b \in \mathcal{B}_{hkr}} \sum_{v \in \mathcal{V}_{hkr}^b} \frac{1}{|\mathcal{V}_{hkr}^b|} \sum_{a \in \mathcal{O}_{hkr}^v} \mathbb{1}_{\{\mathrm{Full}_v^a\}} \cdot W_v^a \tag{B.2}$$

Appendix C. Path-passenger matching

After obtaining the optimal path shares p_{hkr}^*, the operator may need to know which path to provide to each such that the final path shares are close to p_{hkr}^*, especially when passengers have different preferences and may not follow the unpreferred recommendations. In this section, we define a path-passenger matching problem as a solution for this challenge.

Consider a passenger j with a path set $\mathcal{R}_i$. His/her inherent preference (utility) of using path $r \in \mathcal{R}_i$ is denoted as V_i^r. If path r' was recommended, the impact of the recommendation on the utility of path r is denoted as $I_{j,r'}^r$. Hence, his/her overall utility of using path r can be represented as

$$U_j^r = V_j^r + \sum_{r' \in \mathcal{R}_j} x_{i,r'} \cdot I_{j,r'}^r + \xi_j^r \quad \forall r \in \mathcal{R}_i, j \in \mathcal{J} \tag{C.1}$$

where ξ_j^r is the random error. $\mathcal{J}$ is the set of all passengers that need recommendations. $x_{j,r'} = 1$ if passenger j is recommended path r', otherwise $x_{j,r'} = 0$. Let $\pi_{i,r'}^r$ be the conditional probability that passenger j chooses path r given that the recommended path is r'. Assuming a utility-maximizing behavior, we have

$$\pi_{j,r'}^r = \mathbb{P}(V_j^r + I_{j,r'}^r + \xi_j^r \geq V_j^{r''} + I_{j,r'}^{r''} + \xi_j^{r''},\ \forall r'' \in \mathcal{R}_j) \tag{C.2}$$

Different assumptions for the distribution of ξ_j^r can lead to different expressions. For example, if ξ_p^r are i.i.d. Gumbel distributed, the choice probability reduces to multinomial logit model (Train, 2009) and we have

$$\pi_{j,r'}^r = \frac{\exp(V_j^r + I_{j,r'}^r)}{\sum_{r'' \in \mathcal{R}_j} \exp(V_j^{r''} + I_{j,r'}^{r''})} \tag{C.3}$$

The value of V_j^r and $I_{j,r'}^r$ can be calibrated using data from an individual-level survey or smart card, which deserves separate research. For those without such information, this information can be approximated by the population average. When developing the path-passenger matching formulation, we assume $\pi_{j,r'}^r$ is known. Figure C.14 shows an example for the conditional probability matrix. The specific values assume that paths with recommendations are more likely to be chosen.

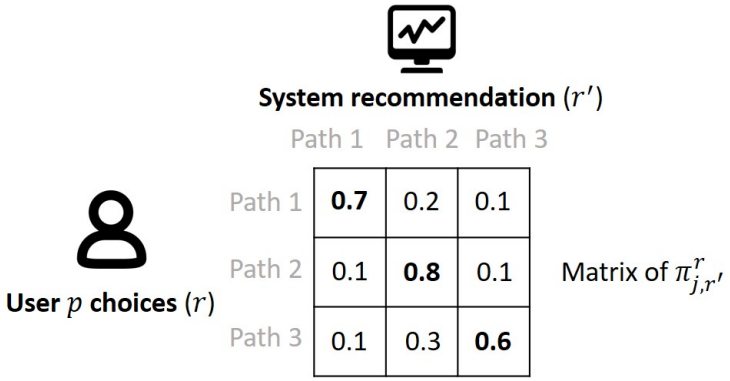


Figure C.14: Example of conditional path choice probability

The conditional probability $\pi^r_{j,r'}$ captures the individual's inherent preference for different paths as well as the response to the recommendation system.

The expected path flow for (h, k, r) is

$$\mu_{hkr}(\boldsymbol{x}) = \sum_{j \in \mathcal{J}_{hk}} \sum_{r' \in \mathcal{R}_k} x_{j,r'} \cdot \pi^r_{j,r'} + q_{hkr} \quad \forall (h, k, r) \in \mathcal{F} \tag{C.4}$$

where $\boldsymbol{x} =: (x_{j,r})_{j \in \mathcal{J}, r \in \mathcal{R}_j}$. q_{hkr} (constant) is the flow of passengers in (h, k, r) that do not need recommendations. $\mathcal{J}_{hk} \subseteq \mathcal{J}$ is the set of passengers with OD pair k and departure time h who need recommendations.

Suppose the value of $\pi^r_{p,r'}$ is known, we can formulate the path-passenger matching problem as an integer linear program:

$$\min_{\boldsymbol{x}} \quad \sum_{(h,k,r) \in \mathcal{F}} |\mu_{hkr}(\boldsymbol{x}) - d_{hk} \cdot p^*_{hkr}| \tag{C.5a}$$

$$\text{s.t.} \quad \text{Constraint } (C.4) \tag{C.5b}$$

$$d_{hk} = \sum_{r \in \mathcal{R}_k} \mu_{hkr}(\boldsymbol{x}) \quad \forall h \in \mathcal{H}, k \in \mathcal{K} \tag{C.5c}$$

$$\sum_{r \in R_j} x_{j,r} = 1 \quad \forall j \in \mathcal{J} \tag{C.5d}$$

$$x_{j,r} \in \{0, 1\} \quad \forall j \in \mathcal{J}, r \in \mathcal{R}_j \tag{C.5e}$$

The objective function aims to minimize the difference between the expected path flow and the optimal path flow. Solving Eq. C.5 yields which path should be recommended to each passenger. It is worth noting that one could also solve the path-passenger matching problem and the path recommendation problem simultaneously, which is equivalent to an individual-based path recommendation problem (Mo et al., 2023).

Appendix D. Validation of the simulation model

Usually, a transit simulator is validated by "OD exit flow" (i.e., the number of tap-out passengers for a specific OD pair and time interval). This is because we only input the "OD entry flow" (i.e., the number of tap-in passengers for a specific OD pair and time interval). Since the simulator will output the tap-out time for each passenger, comparing the model-output OD exit flow with the ground truth (obtained by AFC data with tap-out information) provides validation for the model.

However, in this study, the CTA system does not have tap-out information because it is an open system, implying that the ground truth OD exit flow is not available. But CTA adopted a destination system called "ODX". The ODX algorithm is developed by Sánchez-Martínez (2017). It is shorthand for "origin, destination, and transfer inference algorithm," an extension of the O-D inference algorithm proposed by Zhao et al. (2007). It takes automatically collected data, including AVL and automatic fare collection (AFC), as inputs and infers both destinations and transfers in a tap-on-only transit system, including locations and times. Given a series of tap-in records for a given smart card ID, the tap-out information is inferred as follows: 1) if the current tap-in time is close to the previous tap-in time, the current vehicle "stage" is part of a transfer journey from the previous stage and the alighting location of the previous stage is inferred as the closest stop on that route to the boarding location of the current (second) stage; 2) if there is a large time gap between the current tap-on and the previous tap-int, the alighting location of the previous journey is inferred as the closest stop on the previous route to the boarding location of the current journey assuming passengers' travel patterns are symmetrical and the distance between the inferred alighting location and the subsequent boarding location meets maximum distance criteria. More details can be found in Zhao et al. (2007) and Sánchez-Martínez (2017).

We can treat OD exit flow output by ODX as the ground truth. The comparison is based on the data on a normal weekday without incident. We also aggregated the flows by destinations for better visualization. Figure D.15 shows the comparison between OD exit flow between 9:00-10:00 AM at the top 10 stations in the analysis area (see Figure 5). The flows between simulation and ground truth are consistent, implying that the simulation can well capture the passenger and vehicle dynamics. A more comprehensive validation of the simulation model can be found in Mo et al. (2020), where the case study is based on the Hong Kong Mass Transit Railway system.

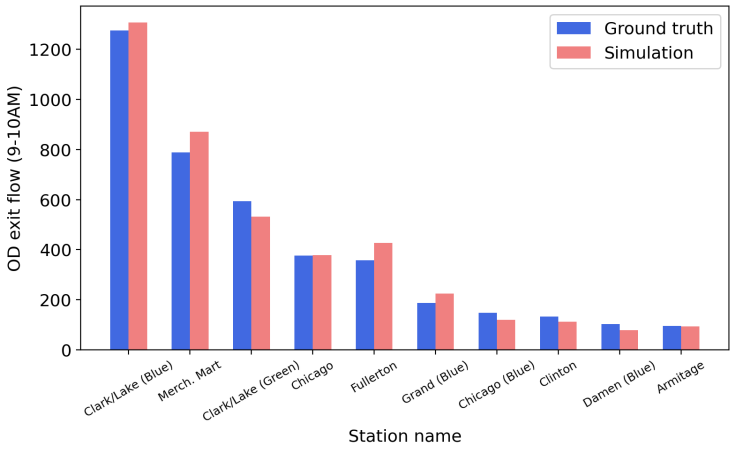


Figure D.15: Comparison between simulation of ground truth OD exit flows

Appendix E. Discussions on the model performance compared to the capacity-based path recommendations

The improvement of the proposed method compared to the capacity-based method mainly comes from the "network-level" optimization for solving the system optimal flows. The proposed model considers the dynamics between upstream and downstream decisions, which are ignored in the capacity-based path recommendations. For example, in the capacity-based path recommendation, when calculating the available

capacity for downstream passengers, the possibly occupied capacity by the recommended upstream passengers is not captured. This is because capacity-based path recommendation is a simple heuristic and does not consider the interaction between the recommendations at differentiation stations from the spatial and temporal aspects. However, the optimization-based methods (both nominal and RO models) are able to capture these interactions through network-level optimization.

In our case study, the improvement compared to the capacity-based method is not significant (around 5% for incident line passengers, see Table 2). The reason is that there are not many upstream and downstream interactions in the Blue Line case study (see Figure 5). Only the parallel bus line has this problem (i.e., the recommendation from upstream passengers may occupy the capacity of downstream passengers). The NS and WE buses have independent lines to connect to the Green and Brown Lines, respectively. And Green and Brown lines have enough capacity to serve passengers from the Blue Line. Therefore, we did not see significant improvement.

This can be further evidenced by the path shares comparison (Figure E.16). The path shares shown in the figure are the weighted average over all OD pairs and time intervals with weights equal to the corresponding demand. We observe that the optimization model de-prioritizes the use of parallel buses. However, the capacity-based model, since cannot capture the upstream and downstream recommendation interactions, over-recommends passengers to the parallel buses. It is worth noting that, when there are no upstream and downstream interactions, the capacity-based path recommendation can be very close to the "system-optimal" path shares (if the travel times of alternative paths are similar).

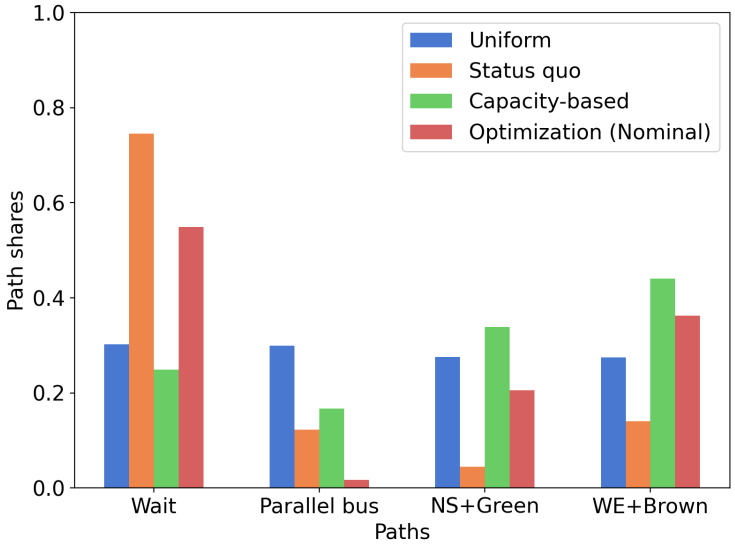


Figure E.16: Path shares comparison for different recommendation strategies

References

Abdelgawad, H., Abdulhai, B., 2012. Large-scale evacuation using subway and bus transit: approach and application in city of toronto. Journal of Transportation Engineering 138, 1215–1232.

Ben-Tal, A., Den Hertog, D., Vial, J.P., 2015. Deriving robust counterparts of nonlinear uncertain inequalities. Mathematical programming 149, 265–299.

Ben-Tal, A., El Ghaoui, L., Nemirovski, A., 2009. Robust Optimization. Princeton Series in Applied Mathematics, Princeton University Press.

Ben-Tal, A., Nemirovski, A., 1998. Robust convex optimization. Mathematics of Operations Research 23, 769–805.

Ben-Tal, A., Nemirovski, A., 1999. Robust solutions of uncertain linear programs. Operations Research Letters 25, 1–13.

Bertsimas, D., Brown, D.B., Caramanis, C., 2011. Theory and applications of robust optimization. SIAM Review 53, 464–501.

Bertsimas, D., den Hertog, D., 2020. Robust and adaptive optimization. Dynamic Ideas LLC, Belmont, Massachusetts.

Bertsimas, D., Litvinov, E., Sun, X.A., Zhao, J., Zheng, T., 2012. Adaptive robust optimization for the security constrained unit commitment problem. IEEE transactions on power systems 28, 52–63.

Bertsimas, D., Sim, M., 2004. The price of robustness. Operations Research 52, 35–53.

Böhmová, K., Mihalák, M., Pröger, T., Srámek, R., Widmayer, P., 2013. Robust routing in urban public transportation: How to find reliable journeys based on past observations, in: ATMOS-13th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems-2013, Schloss Dagstuhl—Leibniz-Zentrum fuer Informatik. pp. 27–41.

Bruglieri, M., Bruschi, F., Colorni, A., Luè, A., Nocerino, R., Rana, V., 2015. A real-time information system for public transport in case of delays and service disruptions. Transportation Research Procedia 10, 493–502.

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J., 2014. An overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research Part B: Methodological 63, 15–37.

Cain, M.K., Zhang, Z., Yuan, K.H., 2017. Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. Behavior research methods 49, 1716–1735.

Cardin, M.A., Deng, Y., Sun, C., 2017. Real options and flexibility analysis in design and management of one-way mobility on-demand systems using decision rules. Transportation Research Part C: Emerging Technologies 84, 265–287.

Chen, S., Di, Y., Liu, S., Wang, B., 2017. Modelling and analysis on emergency evacuation from metro stations. Mathematical Problems in Engineering 2017.

Chen, X.M., Xiong, C., He, X., Zhu, Z., Zhang, L., 2016. Time-of-day vehicle mileage fees for congestion mitigation and revenue generation: A simulation-based optimization method and its real-world application. Transportation Research Part C: Emerging Technologies 63, 71–95.

Chong, L., Osorio, C., 2018. A simulation-based optimization algorithm for dynamic large-scale urban transportation problems. Transportation Science 52, 637–656.

Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M., 2014. Dispatching and coordination in multi-area railway traffic management. Computers & Operations Research 44, 146–160.

Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M., 2010. A tabu search algorithm for rerouting trains during rail operations. Transportation Research Part B: Methodological 44, 175–192.

Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M., 2012. Bi-objective conflict detection and resolution in railway traffic management. Transportation Research Part C: Emerging Technologies 20, 79–94.

Cox, A., Prager, F., Rose, A., 2011. Transportation security and the role of resilience: A foundation for operational metrics. Transport policy 18, 307–317.

D'Ariano, A., Corman, F., Pacciarelli, D., Pranzo, M., 2008. Reordering and local rerouting strategies to manage train traffic in real time. Transportation science 42, 405–419.

De Cea, J., Fernández, E., 1993. Transit assignment for congested public transport systems: an equilibrium model. Transportation science 27, 133–147.

Dollevoet, T., Huisman, D., Schmidt, M., Schöbel, A., 2012. Delay management with rerouting of passengers. Transportation science 46, 74–89.

Dong, J.X., Song, D.P., 2009. Container fleet sizing and empty repositioning in liner shipping systems. Transportation Research Part E: Logistics and Transportation Review 45, 860–877.

D'Ariano, A., Pranzo, M., 2009. An advanced real-time train dispatching system for minimizing the propagation of delays in a dispatching area under severe disturbances. Networks and Spatial Economics 9, 63–84.

Gao, Y., Kroon, L., Schmidt, M., Yang, L., 2016. Rescheduling a metro line in an over-crowded situation after disruptions. Transportation Research Part B: Methodological 93, 425–449.

Guo, X., Caros, N.S., Zhao, J., 2021. Robust matching-integrated vehicle rebalancing in ride-hailing system with uncertain demand. Transportation Research Part B: Methodological 150, 161–189.

Hamdouch, Y., Lawphongpanich, S., 2008. Schedule-based transit assignment model with travel strategies and capacity constraints. Transportation Research Part B: Methodological 42, 663–684.

Hamdouch, Y., Szeto, W., Jiang, Y., 2014. A new schedule-based transit assignment model with travel strategies and supply uncertainties. Transportation Research Part B: Methodological 67, 35–67.

Hassannayebi, E., Memarpour, M., Mardani, S., Shakibayifar, M., Bakhshayeshi, I., Espahbod, S., 2020. A hybrid simulation model of passenger emergency evacuation under disruption scenarios: A case study of a large transfer railway station. Journal of Simulation 14, 204–228.

He, X., Chen, X., Xiong, C., Zhu, Z., Zhang, L., 2017. Optimal time-varying pricing for toll roads under multiple objectives: a simulation-based optimization approach. Transportation Science 51, 412–426.

Jespersen-Groth, J., Potthoff, D., Clausen, J., Huisman, D., Kroon, L., Maróti, G., Nielsen, M.N., 2009. Disruption management in passenger railway transportation, in: Robust and online large-scale optimization. Springer, pp. 399–421.

Jin, J.G., Teo, K.M., Odoni, A.R., 2016. Optimizing bus bridging services in response to disruptions of urban transit rail networks. Transportation Science 50, 790–804.

Kepaptsoglou, K., Karlaftis, M.G., 2009. The bus bridging problem in metro operations: conceptual framework, models and algorithms. Public Transport 1, 275–297.

Kroon, L., Maróti, G., Nielsen, L., 2015. Rescheduling of railway rolling stock with dynamic passenger flows. Transportation Science 49, 165–184.

Lawphongpanich, S., Yin, Y., 2010. Solving the pareto-improving toll problem via manifold suboptimization. Transportation Research Part C: Emerging Technologies 18, 234–246.

Lin, T., Shalaby, A., Miller, E., 2016. Transit user behaviour in response to service disruption: State of knowledge, in: Canadian Transportation Research Forum 51st Annual Conference-North American Transport Challenges in an Era of Change//Les défis des transports en Amérique du Nord à une aire de changement Toronto, Ontario.

Louwerse, I., Huisman, D., 2014. Adjusting a railway timetable in case of partial or complete blockades. European journal of operational research 235, 583–593.

Ma, C., Hao, W., He, R., Jia, X., Pan, F., Fan, J., Xiong, R., 2018. Distribution path robust optimization of electric vehicle with multiple distribution centers. PloS One 13.

Mo, B., Koutsopoulos, H.N., Shen, Z.J.M., Zhao, J., 2023. Individual path recommendation under public transit service disruptions considering behavior uncertainty. arXiv preprint arXiv:2301.00916 .

Mo, B., Ma, Z., Koutsopoulos, H.N., Zhao, J., 2020. Capacity-constrained network performance model for urban rail systems. Transportation Research Record , 0361198120914309.

Mo, B., Ma, Z., Koutsopoulos, H.N., Zhao, J., 2021. Calibrating path choices and train capacities for urban rail transit simulation models using smart card and train movement data. Journal of Advanced Transportation 2021.

Mo, B., Ma, Z., Koutsopoulos, H.N., Zhao, J., 2022a. Ex post path choice estimation for urban rail systems using smart card data: An aggregated time-space hypernetwork approach. Transportation Science .

Mo, B., Von Franque, M.Y., Koutsopoulos, H.N., Attanucci, J.P., Zhao, J., 2022b. Impact of unplanned long-term service disruptions on urban public transit systems. IEEE Open Journal of Intelligent Transportation Systems 3, 551–569. doi:10.1109/OJITS.2022.3199108.

Nguyen, S., Pallottino, S., Malucelli, F., 2001. A modeling framework for passenger assignment on a transport network with timetables. Transportation Science 35, 238–249.

Nielsen, O.A., 2000. A stochastic transit assignment model considering differences in passengers utility functions. Transportation Research Part B: Methodological 34, 377–402.

Niu, H., Zhou, X., 2013. Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. Transportation Research Part C: Emerging Technologies 36, 212–230.

Noordhoek, M., Dullaert, W., Lai, D.S., de Leeuw, S., 2018. A simulation–optimization approach for a service-constrained multi-echelon distribution network. Transportation Research Part E: Logistics and Transportation Review 114, 292–311.

Osorio, C., Bierlaire, M., 2013. A simulation-based optimization framework for urban transportation problems. Operations Research 61, 1333–1345.

Osorio, C., Nanduri, K., 2015a. Energy-efficient urban traffic management: a microscopic simulation-based approach. Transportation Science 49, 637–651.

Osorio, C., Nanduri, K., 2015b. Urban transportation emissions mitigation: Coupling high-resolution vehicular emissions and traffic models for traffic signal optimization. Transportation Research Part B: Methodological 81, 520–538.

Rahimi, E., Shamshiripour, A., Shabanpour, R., Mohammadian, A., Auld, J., 2019. Analysis of transit users' waiting tolerance in response to unplanned service disruptions. Transportation Research Part D: Transport and Environment 77, 639–653.

Rahimi, E., Shamshiripour, A., Shabanpour, R., Mohammadian, A., Auld, J., 2020. Analysis of transit users' response behavior in case of unplanned service disruptions. Transportation Research Record , 0361198120911921.

Roelofsen, D., Cats, O., van Oort, N., Hoogendoorn, S., 2018. Assessing disruption management strategies in rail-bound urban public transport systems from a passenger perspective, in: Proceedings of the 14th Conference on Advanced Systems in Public Transport (CASPT), Brisbane, Australia.

Sánchez-Martínez, G.E., 2017. Inference of public transportation trip destinations by using fare transaction and vehicle location data: Dynamic programming approach. Transportation Research Record 2652, 1–7.

Schachtebeck, M., Schöbel, A., 2010. To wait or not to wait—and who goes first? delay management with priority decisions. Transportation science 44, 307–321.

Schmöcker, J.D., Bell, M.G., Kurauchi, F., 2008. A quasi-dynamic capacity constrained frequency-based transit assignment model. Transportation Research Part B: Methodological 42, 925–945.

Schmöcker, J.D., Fonzone, A., Shimamoto, H., Kurauchi, F., Bell, M.G., 2011. Frequency-based transit assignment considering seat capacities. Transportation Research Part B: Methodological 45, 392–408.

Schöbel, A., 2007. Integer programming approaches for solving the delay management problem, in: Algorithmic methods for railway optimization. Springer, pp. 145–170.

Shi, X., Li, X., 2021. Operations design of modular vehicles on an oversaturated corridor with first-in, first-out passenger queueing. Transportation Science 55, 1187–1205.

Spall, J.C., 1997. A one-measurement form of simultaneous perturbation stochastic approximation. Automatica 33, 109–112.

Tan, Z., Xu, M., Meng, Q., Li, Z.C., 2020. Evacuating metro passengers via the urban bus system under uncertain disruption recovery time and heterogeneous risk-taking behaviour. Transportation research part C: emerging technologies 119, 102761.

Train, K.E., 2009. Discrete choice methods with simulation. Cambridge university press.

Wang, J., Yuan, Z., Yin, Y., 2019a. Optimization of bus bridging service under unexpected metro disruptions with dynamic passenger flows. Journal of Advanced Transportation 2019.

Wang, X., Chen, S., Zhou, Y., Peng, H., Cui, Y., 2013. Simulation on passenger evacuation under fire emergency in metro station, in: 2013 IEEE International Conference on Intelligent Rail Transportation Proceedings, IEEE. pp. 259–262.

Wang, Y., Ning, B., Tang, T., Van Den Boom, T.J., De Schutter, B., 2015a. Efficient real-time train scheduling for urban rail transit systems using iterative convex programming. IEEE Transactions on Intelligent Transportation Systems 16, 3337–3352.

Wang, Y., Tang, T., Ning, B., Van Den Boom, T.J., De Schutter, B., 2015b. Passenger-demands-oriented train scheduling for an urban rail transit network. Transportation Research Part C: Emerging Technologies 60, 1–23.

Wang, Y., Zhang, Y., Tang, J., 2019b. A distributionally robust optimization approach for surgery block allocation. European Journal of Operational Research 273, 740–753.

Wu, J.H., Florian, M., Marcotte, P., 1994. Transit equilibrium assignment: a model and solution algorithms. Transportation Science 28, 193–203.

Xiong, P., Jirutitijaroen, P., Singh, C., 2016. A distributionally robust optimization model for unit commitment considering uncertain wind power generation. IEEE Transactions on Power Systems 32, 39–49.

Yang, X., Li, X., Gao, Z., Wang, H., Tang, T., 2012. A cooperative scheduling model for timetable optimization in subway systems. IEEE Transactions on Intelligent Transportation Systems 14, 438–447.

Yin, J., Tang, T., Yang, L., Gao, Z., Ran, B., 2016. Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: An approximate dynamic programming approach. Transportation Research Part B: Methodological 91, 178–210.

Zhan, S., Kroon, L.G., Veelenturf, L.P., Wagenaar, J.C., 2015. Real-time high-speed train rescheduling in case of a complete blockage. Transportation Research Part B: Methodological 78, 182–201.

Zhang, W., et al., 2017. Simulation-based robust optimization for the schedule of single-direction bus transit route: The design of experiment. Transportation Research Part E: Logistics and Transportation Review 106, 203–230.

Zhao, J., Rahbee, A., Wilson, N.H., 2007. Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. Computer-Aided Civil and Infrastructure Engineering 22, 376–387.

Zhou, M., Dong, H., Zhao, Y., Ioannou, P.A., Wang, F.Y., 2019. Optimization of crowd evacuation with leaders in urban rail transit stations. IEEE transactions on intelligent transportation systems 20, 4476–4487.