Distributionally Robust Fair Transit Resource Allocation During a Pandemic

Luying Sun
Department of Industrial & Systems Engineering, Virginia Tech, Blacksburg, VA 24061, luyingsun@vt.edu.

Weijun Xie
Department of Industrial & Systems Engineering, Virginia Tech, Blacksburg, VA 24061, wxie@vt.edu.

Tim Witten
Blacksburg Transit, Blacksburg, VA 24060, TWitten@blacksburg.gov.

This paper studies Distributionally robust Fair transit Resource Allocation model (DrFRAM) under the Wasserstein ambiguity set to optimize the public transit resource allocation during a pandemic. We show that the proposed DrFRAM is highly nonconvex and nonlinear and is NP-hard in general. Fortunately, we show that DrFRAM can be reformulated as a mixed-integer linear programming (MILP) by leveraging the equivalent representation of distributionally robust optimization and monotonicity properties, binarizing integer variables, and linearizing nonconvex terms. To improve the proposed MILP formulation, we derive stronger ones and develop valid inequalities by exploiting the model structures. Besides, we develop scenario decomposition methods using different MILP formulations to solve the scenario subproblems and introduce a simple yet effective No-one-left based approximation algorithm with a provable approximation guarantee to solve the model to near optimality. Finally, we numerically demonstrate the effectiveness of the proposed approaches and apply them to real-world data provided by the Blacksburg Transit.

Key words: Distributionally Robust, Mixed-Integer Programming, Strong Formulations, Valid Inequalities
History:

1. Introduction

The recent outbreak of the SARS-CoV-2 (i.e., COVID-19 disease) has been profoundly influencing society. Many schools have been shut down and turned to online; many restaurants and shopping malls have either been closed or significantly reduced capacity; many firms have required employees to work from home. On the other hand, as the vaccination process has been accelerated and the promise of the vaccines, the number of COVID-19 infection cases in the U.S. has been decreasing since January 2021. Therefore, many states are now re-opening the business as well as schools; for example, the Commonwealth of Virginia's public schools have been re-opened since March 2021. Starting from August 2021, as the viruses are constantly evolving into new variants which may be more infectious than the original ones, many states restart the mask mandates in public conveyances and indoor activities. Undoubtedly, these policies have tremendously influenced people's daily activities, which causes a significant demand uncertainty for public transit. Besides demand

uncertainty, to keep passengers safe and prevent transmission of the coronaviruses, it is of vital importance to observe the social distancing within the public transit, which in turn will reduce the capacity of transit vehicles (e.g., buses) and thus affect the entire operations of the public transit systems. This paper is trying to solve this issue. The results in this paper are also useful for another wave of COVID-19 outbreaks or preparing for a future pandemic.

The authors collaborate with local Blacksburg Transit, a daily outdoor transportation tool used by many students and residents at the Town of Blacksburg. Since the training of employees and installation of bus stops can cause much longer delays and aggravate the under-staffing issue, the transit operators in Blacksburg Transit would like to keep the current routes and stops as they are but to optimize the transit resources, i.e., bus allocation, to maximize their utilization rates as well as social equity. This motivates us to study the fair transit resource allocation problem. As Blacksburg Transit shares us the passenger alighting and boarding data, we are able to build a data-driven optimization model to leverage these available data to account for future uncertainties.

1.1. Literature Reviews

Transit system design and optimization have been widely studied in the literature (see, e.g., Daganzo and Ouyang 2010, Ceder 2016, and references therein). The works in transit are mainly on optimal scheduling and routing to serve the public better. For example, Chien and Schonfeld (1997) incorporated spatial characteristics and demand patterns of urban areas into their optimization model. Other important earlier works can be found in De Cea and Fernández (1993), Quadrifoglio et al. (2006), Guan et al. (2006), Jeihani et al. (2013) or in the review papers (Guihaire and Hao 2008, Hörcher and Tirachini 2021). Recently, Nourbakhsh and Ouyang (2012) focused on the flexible transit design in the low-demand areas, and Ouyang et al. (2014) proposed a continuum approximation approach for bus design under heterogeneous demand. Chen and Nie (2017) attempted to integrate an e-hailing system with public transit. Iliopoulou and Kepaptsoglou (2019) jointly optimized network design and charging infrastructure locations for the electric buses. Abdolmaleki et al. (2020) built a new model for synchronizing timetables in a transit network that minimizes the total transfer waiting time. Other interesting works can be found in Román-De la Sancha et al. (2018), Matisziw et al. (2006), Jin et al. (2016), Agarwal and Ergun (2008), Mahéo et al. (2019), among many others. Different from these works, we focus on transit resource allocation given the predetermined routes.

Recent advances in stochastic and robust optimization provide tools for the transit design under uncertainty (see, e.g., Li et al. 2009, Fernandes et al. 2018, Li et al. 2008, Fu and Lam 2014, Hamdouch et al. 2014, Hadas and Shnaiderman 2012 for the earlier interesting works). For example, Kulshrestha et al. (2014) developed a robust model for bus dispatching for evacuations to optimize

the total travel time under demand uncertainty. An and Lo (2016) studied transit system design under demand uncertainty using two-stage stochastic programming. Liang et al. (2019) proposed a two-step modeling framework for the bus transit network design considering an existing metro network and demand uncertainty. In Yoon and Chow (2020), the authors provided a sequential learning framework to address the demand uncertainty when designing lines and routes in the transit system. Shehadeh (2020) studied two distributionally robust models to address the randomness in demands for a mobile facility routing and scheduling problem. Different from existing ones, our modeling framework is data-driven without any knowledge of the underlying distribution, which uses real-world data from a transit system. Besides, to fulfill the requirements of transit operators, our proposed model does not alter the existing lines and routes.

The study of transit design during a pandemic, especially the COVID-19 pandemic, is rather limited. As shown in Liu et al. (2020), the COVID-19 pandemic caused a major transit demand decline for many public transit systems in the United States due to the fear of contracting the disease, practicing social distancing, and lockdown policy, which has also been observed by many social workers. Mo et al. (2021) studied the epidemic spreading model in the time-varying transit network. Chen et al. (2020) redesigned the campus bus systems by shortening the routes and enforcing social distancing. The most relevant to our research are Gkiotsalitis and Cats (2021) and Yang and Nie (2020), which optimized metro service frequency during the COVID-19 pandemic. Different from Gkiotsalitis and Cats (2021), and Yang and Nie (2020), we not only optimize the number of buses assigned to different routes but also try to stick to one type of bus per route to avoid drivers' unnecessary confusion. More importantly, we incorporate both demand uncertainty and alighting rate uncertainty, and thus our proposed model is more flexible and data-driven. The scope of our model is also different from Gkiotsalitis and Cats (2021), Chen et al. (2020). Instead of minimizing the operation costs or maximizing the profits, we focus on optimizing social equity for each route and minimizing the passenger abandon rate due to capacity restrictions.

1.2. Summary of Main Contributions

The objective of this study, motivated by Blacksburg Transit, is to determine optimal transit resource allocation to minimize the highest bus utilization rate and the largest passenger abandon rate with limited bus operation data given to achieve better social equity under stochastic passenger arrival and alighting rates and overcome the future uncertainties as much as we can to serve the passengers better. This gives rise to Distributionally robust Fair transit Resource Allocation model (DrFRAM) under a data-driven ambiguity set. The main contributions of DrFRAM are summarized as below:

(i) We study DrFRAM under type-∞ Wasserstein ambiguity set. As far as we are concerned, our DrFRAM model is the first one focusing on fair multi-type transit resource allocation under passenger arrival and alighting rates uncertainty. We prove that DrFRAM cannot be solved in polynomial time unless $P = NP$ even when the problem is deterministic, and there are only two types of buses.

(ii) We derive the monotonicity properties of the DrFRAM, using which we can significantly simplify the DrFRAM and linearize the nonlinear components in the objective of DrFRAM.

(iii) We propose a mixed-integer linear programming (MILP) formulation for DrFRAM by binarizing integer variables and linearizing the nonlinear functions using McCormick inequalities. To strengthen the MILP formulation, we derive stronger ones and develop valid inequalities by exploiting the model structures. These formulations allow us to design efficient scenario decomposition methods.

(iv) Besides these exact methods, we develop an approximate scenario decomposition method using objective cuts to speed up the convergence and No-one-left based approximation algorithm for solving DrFRAM to near-optimality. Both approaches come with theoretical approximation guarantees, demonstrating the strengths of these approaches.

(v) We numerically demonstrate the effectiveness of the proposed approaches, outperforming the sampling average approximation method, and apply them to solve the real-world instances using the data provided by Blacksburg Transit.

Notation. The following notation is used throughout the paper. We use bold letters (e.g., $\boldsymbol{x}, \boldsymbol{A}$) to denote vectors and matrices and use corresponding non-bold letters to denote their components. Given a vector or matrix $\boldsymbol{x}$, its zero norm $\|\boldsymbol{x}\|_0$ denotes the number of its nonzero elements. We let $\mathbf{e}$ be the vector or matrix of all ones and let $\mathbf{e}_i$ be the ith standard basis vector. Given an integer n, we let $[n] := \{1, 2, \ldots, n\}$, and use $\mathbb{R}^n_+ := \{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0, \forall i \in [n]\}$. Given a real number t, we let $(t)_+ := \max\{t, 0\}$. Given a finite set I, we let $|I|$ denote its cardinality. We let $\tilde{\boldsymbol{\xi}}$ denote a random vector and denote its realizations by $\boldsymbol{\xi}$. We use superscript $k \in [N]$ to denote the index of scenario k. For a matrix $\boldsymbol{A}$, we let $\boldsymbol{A}_{i:}$ denote ith row of $\boldsymbol{A}$ and $\boldsymbol{A}_{:j}$ denote jth column of $\boldsymbol{A}$. Additional notations will be introduced as needed.

Organization. The remainder of the paper is organized as follows. In Section 2, we introduce the model formulation. Sections 3 and 4 show model properties and derive an equivalent mixed-integer linear programming formulation. Section 5 shows the convexification of two substructures and derivation of strong valid inequalities, and Section 6 develops approximation algorithms. Finally, Section 7 numerically demonstrates the effectiveness of the proposed solution approaches, and Section 8 concludes this paper.

2. Model Formulation

We collaborate with Blacksburg Transit in this research. Since March 2020, Blacksburg Transit has been making efforts to prevent the spread of the pandemic and reduce the risk of exposure to COVID-19 in order to support the town better. In particular, they are curious about the optimal plan of bus resource allocation to minimize the highest bus utilization rate during the pandemic. They are willing to provide us with recent transit sensor data about passenger's arrival and alighting information to support our research. However, due to various reasons, such as sensor malfunctioning, only a small subset of the data provided to us can be useful in our research. Besides, the transit operators have little knowledge of underlying probability distribution nor moment information of the passenger arrival or alighting rates, especially during the pandemic. Therefore, motivated by Blacksburg Transit, we plan to formulate the problem as a Distributionally robust Fair transit Resource Allocation model (DrFRAM) using a data-driven ambiguity set, which allows us to leverage the collected dataset to its maximum extent and account for non-stationary uncertainties in the near future.

For the strategic planning purpose as well as considering the intrinsic problem structure of DrFRAM that passenger's arrival and alighting data are available, we adopt the modeling of a transit system based on the seminal work in Hadas and Shnaiderman (2012) using aggregate arrival and alighting rates and not relying on the user-specific origin-destination data, since the former is easy to obtain and the latter is difficult to get access to in the transit data available to us. In DrFRAM, there are K different types of buses, where each bears a nominal capacity c_k for each $k \in [K]$, and there are η_k type-k buses. Note that during a pandemic, to observe the social distancing strictly, the bus capacities will be decreased proportionally and strictly enforced to keep passengers safe and sound. Thus, we let a pandemic factor $\delta_k \in [0,1]$ denote the percent of a type k bus capacity that allows being operated. That is, only $\lfloor \delta_k c_k \rfloor$ passengers will be allowed onto the bus during a pandemic. Suppose that there are I distinct bus routes, and for each bus route $i \in [i]$, there are J_i bus stops. Since passengers can alight and depart a bus at any bus stop for each $i \in [I]$ and $j \in [J_i]$, we let the random parameter $\tilde{a}_{ij} \in [0,1]$ denote the proportion of passengers alighting from the bus and let the random variable $\tilde{\lambda}_{ij} \in \mathbb{Z}_+$ represent the passenger arrival rate during a unit time interval, i.e., the number of passengers arriving during that time. Note that $\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{\lambda}}$ are both random since they can vary over time and change with severeness of the pandemic. For notational convenience, we let $\tilde{\boldsymbol{\xi}} = (\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{\lambda}})$. Since the historical data are available to us, in DrFRAM, we use the Wasserstein distance to characterize the random parameters, which can both ensure the model to data-driven and robust (see, e.g., Kuhn et al. 2019).

The transit operators would like to know that given the predetermined routes (so they should not re-train their employees), what is the fairest way to allocate buses so that the passengers will

be the best serviced and the safety policies should be strictly enforced? To help them make a better decision, we let the integer variable $n_{ik} \in \mathbb{Z}_+$ denote the number of type $k \in [K]$ buses allocated to the $i \in [I]$th route. Due to consistency and for ease of management, the transit operators would like to ensure that each route is subject to the same type of bus whenever possible. Thus, we let the binary variable $x_{ik} \in \{0,1\}$ denote whether type k buses will be assigned to route $i \in [I]$ or not. Using this notation, we see that $\sum_{k\in[K]} x_{ik}\lfloor \delta_k c_k \rfloor$ represents the bus capacity at route $i \in [I]$. In our model, given a realization $\boldsymbol{\xi}$ of the random parameters $\tilde{\boldsymbol{\xi}}$, we let $L_{ij}(\boldsymbol{\xi})$ denote the number of passengers remaining on a bus right after the jth stop at route $i \in [I]$. Due to its physical meaning and the fact that the arrival rates can be quite low during the pandemic, we keep the variable $L_{ij}(\boldsymbol{\xi})$ as an integer[1]. To achieve social equity, we define the bus utilization rate as the ratio of the number of passengers on the bus to the bus capacity, and the passenger abandon rate as the ratio of the number of passengers who cannot get on board due to limited bus capacity to the number of passengers arrived. The operation managers would like to know an optimal way to allocate the buses, which is fair to passengers from different routes. That is, they want to ensure that both the highest bus utilization rate and the largest passenger abandon rate due to limited bus capacity among all the routes are low.

In DrFRAM, as passenger arrival and alighting rates are often subject to change over time, and their joint probability distribution is difficult to estimate, these two parameters are supposed to be random. Besides, we assume that

(i) the bus routes are predetermined;
(ii) each route is subject to the same type of bus;
(iii) traffic conditions (for example, congestion and interaction with pedestrians and other vehicles), staffing are not considered in our model; and
(iv) limited passenger arrival and alighting data are available.

After talking to Blacksburg Transit operators, we were informed that these assumptions are consistent with the daily operations in Blacksburg Transit during the COVID-19 pandemic. With the notation introduced above, we are ready to present the mathematical formulation of DrFRAM as below:

$$\text{(DrFRAM)} \qquad v^* = \min_{\boldsymbol{n},\boldsymbol{x}} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\mathcal{Q}\left(\boldsymbol{n},\boldsymbol{x},\tilde{\boldsymbol{\xi}}\right)\right], \tag{1a}$$

$$\text{s.t.} \quad \sum_{k\in[K]} x_{ik} = 1, \forall i \in [I], \tag{1b}$$

$$n_{ik} \le \eta_k x_{ik}, \forall i \in [I], \forall k \in [K], \tag{1c}$$

[1] Our numerical result in Appendix B shows that relaxing the integrality of $L_{ij}(\boldsymbol{\xi})$ can cause a much smaller objective value and may even lead to a misleading conclusion.

$$\sum_{i\in[I]} n_{ik} \le \eta_k, \forall k \in [K], \tag{1d}$$

$$\sum_{k\in[K]} n_{ik} \ge 1, \forall i \in [I], \tag{1e}$$

$$x_{ik} \in \{0,1\}, n_{ik} \in \mathbb{Z}^+, \forall i \in [I], \forall k \in [K]. \tag{1f}$$

In DrFRAM (1), the objective is to minimize the worst-case resource planning outcomes, where ambiguity set $\mathcal{P}$ denotes a family of probability distributions, and $\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \tilde{\boldsymbol{\xi}})$ denotes the random recourse function which will be specified later. Constraints (1b) ensure that each route will commit to one type of bus. Constraints (1c) and (1d) jointly show that the number of buses allocated to a particular route is no larger than the number of available buses. Constraints (1e) show that each route should have at least one bus to achieve social equity. Constraints (1f) specify the boundaries of the decision variables.

Given a realization $\boldsymbol{\xi}$ of random parameters $\tilde{\boldsymbol{\xi}}$ and the values of first-stage decisions $(\boldsymbol{n}, \boldsymbol{x})$, we can express the recourse function in the following way:

$$\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi}) = \min_{\boldsymbol{L}(\boldsymbol{\xi})} Q(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{L}(\boldsymbol{\xi})) := \left\{ \max_{i\in[I], j\in[J_i]} \underbrace{\frac{L_{ij}(\boldsymbol{\xi})}{\sum_{k\in[K]} n_{ik} \lfloor \delta_k c_k \rfloor}}_{\text{bus utilization rate}} \right.$$

$$\left. + \omega \max_{i\in[I], j\in[J_i]} \underbrace{\max\left[0, 1 + \frac{1}{\lambda_{ij}} \left(\lceil (1-a_{ij}) L_{i,j-1}(\boldsymbol{\xi}) \rceil - \sum_{k\in[K]} n_{ik} \lfloor \delta_k c_k \rfloor \right) \right]}_{\text{passenger abandon rate}} \right\}, \tag{2a}$$

$$\text{s.t.} \quad L_{ij}(\boldsymbol{\xi}) = \min \left\{ \sum_{k\in[K]} n_{ik} \lfloor \delta_k c_k \rfloor, \lceil (1-a_{ij}) L_{i,j-1}(\boldsymbol{\xi}) \rceil + \lambda_{ij} \right\}, \forall i \in [I], \forall j \in [J_i], \tag{2b}$$

$$L_{i0}(\boldsymbol{\xi}) = 0, L_{ij}(\boldsymbol{\xi}) \in \mathbb{Z}^+, \forall i \in [I], \forall j \in [J_i], \tag{2c}$$

where $\omega \ge 0$ is the weight that balances the importance of the highest bus utilization rate and the largest passenger abandon rate. The objective (2a) is to minimize the weighted highest bus utilization rate and the largest passenger abandon rate to enhance the social equity. Note that we choose the min-max fairness measure in (2), widely used in the fairness-related literature (Radunovic and Le Boudec 2007, Du et al. 2017). Constraints (2b) postulate the number of passengers on board at each stop, which follows the convention from the existing transit literature (see, e.g., Hadas and Shnaiderman 2012). Constraints (2c) specify the boundary conditions of the recourse decisions, i.e., the initialization and integrality of variables $\boldsymbol{L}(\boldsymbol{\xi})$.

As mentioned before, the data from Blacksburg Transit has no moment information, and limited data available is not likely for us to estimate the moments accurately. Thus, we decide not to use

a moment-based ambiguity set in this research. Instead, to fully exploit the limited data, we plan to focus on the more appropriate Wasserstein ambiguity set due to its consistency, tractability, computational advantages, as well as its flexibility of accounting for non-stationary uncertainties in the near future.

2.1. Wasserstein Ambiguity Set

In this subsection, we briefly introduce the notion of the Wasserstein ambiguity set and its attractive properties, which are suitable for DrFRAM. In this work, we were provided historical data collected during the operations. Thus, given an empirical distribution $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ constructed using i.i.d. historical data $Z = \{\bar{\boldsymbol{\zeta}}^\ell = (\bar{\boldsymbol{a}}^\ell, \bar{\boldsymbol{\lambda}}^\ell)\}_{\ell\in[N]}$ such that $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}\{\tilde{\boldsymbol{\zeta}} = \bar{\boldsymbol{\zeta}}^\ell\} = 1/N$, this paper considers the data-driven Wasserstein ambiguity set (see, e.g., Gao and Kleywegt 2016, Blanchet et al. 2019a, Esfahani and Kuhn 2018, Blanchet and Murthy 2019, Hanasusanto and Kuhn 2018, Chen et al. 2018, Xie 2019, Abadeh et al. 2018, Kuhn et al. 2019, Blanchet et al. 2019b, Chen and Xie 2019) as below:

$$\mathcal{P}_q = \left\{\mathbb{P} : W_q\left(\mathbb{P}, \mathbb{P}_{\tilde{\boldsymbol{\zeta}}}\right) \le \theta\right\}, \tag{3}$$

where $\theta \ge 0$ denotes the Wasserstein radius and for any $q \in [1,\infty]$, the Wasserstein distance $W_q(:,:)$ is defined as

$$W_q\left(\mathbb{P}_1, \mathbb{P}_2\right) = \inf\left\{\sqrt[q]{\int_{\Xi\times\Xi} \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|^q \mathbb{Q}(d\boldsymbol{\xi}_1, d\boldsymbol{\xi}_2)} : \begin{array}{l}\mathbb{Q} \text{ is a joint distribution of } \tilde{\boldsymbol{\xi}}_1 \text{ and } \tilde{\boldsymbol{\xi}}_2 \\ \text{with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively}\end{array}\right\}.$$

where the support

$$\Xi = \{(\boldsymbol{a}, \boldsymbol{\lambda}) : a_{ij} \in [0,1], \lambda_{ij} \in \mathbb{Z}_+, \forall i \in [I], \forall j \in [J_i]\}.$$

When $q = \infty$, it reduces to the $\infty-$Wasserstein distance

$$W_\infty\left(\mathbb{P}_1, \mathbb{P}_2\right) = \inf\left\{\mathbb{Q} - \operatorname{ess\,sup} \|\tilde{\boldsymbol{\xi}}_1 - \tilde{\boldsymbol{\xi}}_2\| : \begin{array}{l}\mathbb{Q} \text{ is a joint distribution of } \tilde{\boldsymbol{\xi}}_1 \text{ and } \tilde{\boldsymbol{\xi}}_2 \\ \text{with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively}\end{array}\right\}.$$

Above, $\mathbb{Q}$-ess sup $\|\cdot\|$ is the essential supreme of $\|\cdot\|$ with respect to the joint distribution $\mathbb{Q}$, which is formally defined as

$$\mathbb{Q} - \operatorname{ess\,sup} \|\tilde{\xi}_1 - \tilde{\xi}_2\| = \inf\left\{\Delta : \mathbb{Q}[\|\tilde{\xi}_1 - \tilde{\xi}_2\| > \Delta] = 0\right\}.$$

It has been shown in Bertsimas et al. (2018) that its corresponding type $\infty-$Wasserstein ambiguity set $\mathcal{P}_\infty$ has the following equivalent form

$$\mathcal{P}_\infty = \left\{\frac{1}{N}\sum_{\ell\in[N]} \chi(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^\ell) : \exists \boldsymbol{\xi}^\ell \in \Xi, \|\boldsymbol{\xi}^\ell - \bar{\boldsymbol{\zeta}}^\ell\| \le \theta, \forall \ell \in [N]\right\}, \tag{4}$$

where $\chi(\cdot)$ is the Dirac delta function. This neat representation has a straightforward interpretation, i.e., the worst-case distribution is also supported by N points, and each support point can only deviate at most θ amount from one of the empirical data $Z = \{\bar{\boldsymbol{\zeta}}^k\}_{k\in[N]} \subset \Xi$.

Since $W_\infty(\cdot,\cdot)$ has a similar statistical performance as $W_q(\cdot,\cdot)$ with $q \in [1,\infty)$ (Fournier and Guillin 2015, Trillos and Slepčev 2015, Xie 2020, Xie et al. 2021) but has more attractive computational properties, we focus on $\infty-$Wasserstein ambiguity set, i.e., throughout this paper, we suppose that $\mathcal{P} = \mathcal{P}_\infty$.

3. Model Properties

This section explores the properties of DrFRAM (1) and shows its computational complexity.

According to the worst-case characterization in (4) of type-∞ Wasserstein ambiguity set $\mathcal{P} = \mathcal{P}_\infty$. DrFRAM (1) can be equivalently represented as the following deterministic counterpart:

$$\text{(DrFRAM)} \qquad v^* = \min_{\boldsymbol{n},\boldsymbol{x}} \left\{ \frac{1}{N} \sum_{\ell\in[N]} \sup_{\boldsymbol{\xi}^\ell \in \Xi, \|\boldsymbol{\xi}^\ell - \boldsymbol{\zeta}^\ell\| \le \theta} \mathcal{Q}\left(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi}^\ell\right) : (1\text{b}) - (1\text{f}) \right\}. \tag{5}$$

Model (5) can be interpreted as a robustification of Sampling Average Approximation (SAA) by choosing the worst-case perturbation for each empirical sample. Our model properties are based on this reformulation, and we finally choose a particular norm $\|\cdot\|$ in Section 3.2 to further simplify DrFRAM (5).

3.1. Computational Complexity of DrFRAM (5)

We observe that DrFRAM (5) is a mixed-integer nonlinear program and thus expect that it is an NP-hard problem. This motivates us to derive mixed-integer programming techniques and approximation algorithms to solve it.

The NP-hardness of DrFRAM (5) is established upon the well-known NP-complete problem – partition problem.

Proposition 1 *Solving DrFRAM (5) is NP-hard even when $N = 1, K = 2, \theta = 0$.*

Proposition 1 shows that even under simple settings, DrFRAM (5) may not be polynomial-time solvable. This motivates us to develop exact and approximation algorithms to solve it, which serve different purposes. For example, exact algorithms can be useful to verify the correctness of approximation algorithms and can solve many instances to optimality, while approximation algorithms, usually more scalable, can provide high-quality initial solutions to speed up the exact ones.

3.2. Theoretical Sensitivity Analyses and Model Simplification

In this subsection, we analyze the properties of the recourse function $\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi})$ and the second-stage objective function $Q(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{L}(\boldsymbol{\xi}))$. These properties allow us to simplify DrFRAM (5) by linearizing the nonlinear objective function.

We begin with the results for the second-stage objective function $Q(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{L}(\boldsymbol{\xi}))$.

Proposition 2 *For the second-stage objective function $Q(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{L}(\boldsymbol{\xi}))$, we have the following sensitivity results:*

- *Function $Q(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{L}(\boldsymbol{\xi}))$ is monotone non-decreasing as the recourse decision $L_{ij}(\boldsymbol{\xi})$ increases for some $i \in [I]$ and $j \in [J_i]$;*
- *Function $Q(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{L}(\boldsymbol{\xi}))$ is monotone non-increasing as the first-stage decision n_{ik} increases for some $i \in [I]$ and $k \in [K]$;*
- *Function $Q(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{L}(\boldsymbol{\xi}))$ is monotone non-decreasing as the passenger arrival rate λ_{ij} increases for some $i \in [I]$ and $j \in [J_i]$; and*
- *Function $Q(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{L}(\boldsymbol{\xi}))$ is monotone non-decreasing as the passenger alighting rate a_{ij} increases for some $i \in [I]$ and $j \in [J_i]$.*

Proof: We prove the results according to their orders.

- The first monotonicity result is simply because when the number of passengers on the buses increases, both bus utilization rate and passenger abandon rate grow or stay the same.
- The second monotonicity result is because when the number of buses allocated to a particular route increases, both its bus utilization rate and passenger abandon rate decrease or stay the same.
- The third and the fourth monotonicity results are because when more passengers arrive at a bus stop of a particular route, the route's bus utilization rate and passenger abandon rate increase or stay the same.

□

Since the minimization operator does not change the monotonicity of a function, results in Parts (ii)-(iv) still hold for the recourse function $\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi})$.

Corollary 1 *For the recourse function $\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi})$, we have the following sensitivity results:*

- *Recourse function $\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi})$ is monotone non-increasing as the first-stage decision n_{ik} increases for some $i \in [I]$ and $k \in [K]$;*
- *Recourse function $\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi})$ is monotone non-decreasing as the passenger arrival rate λ_{ij} increases for some $i \in [I]$ and $j \in [J_i]$; and*
- *Recourse function $\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{\xi})$ is monotone non-decreasing as the passenger alighting rate a_{ij} increases for some $i \in [I]$ and $j \in [J_i]$.*

The results in Corollary 1 inspire us to choose a proper norm $\|\cdot\|$ in the type-∞ Wasserstein ambiguity set. Specifically, since random parameters $\tilde{\boldsymbol{a}}$ and $\tilde{\boldsymbol{\lambda}}$ have different magnitude, we choose the weighted ℓ_∞ norm as $\|\boldsymbol{\xi}\| = \max\{\gamma\|\boldsymbol{a}\|_\infty, \|\boldsymbol{\lambda}\|_\infty\}$ (recall that we let $\boldsymbol{\xi} = (\boldsymbol{a}, \boldsymbol{\lambda})$), where the positive weight $\gamma > 0$ is to demonstrate the balance of the importance of both parameters. In

practice, one can choose a proper γ such that the scaled vector $\gamma\boldsymbol{a}$ and vector $\boldsymbol{\lambda}$ have a similar order of magnitude.

In the spirits of Corollary 1 and the weighted ℓ_∞ norm, DrFRAM (5) admits the following equivalent representation.

Proposition 3 *Suppose* $\|\boldsymbol{\xi}\| = \max\{\gamma\|\boldsymbol{a}\|_\infty, \|\boldsymbol{\lambda}\|_\infty\}$ *for some* $\gamma > 0$*, then we have*

$$\text{(DrFRAM)} \qquad v^* = \min_{\boldsymbol{n},\boldsymbol{x}} \left\{ \frac{1}{N} \sum_{\ell\in[N]} \mathcal{Q}\left(\boldsymbol{n}, \boldsymbol{x}, \widehat{\boldsymbol{\xi}}^\ell\right) : (1\text{b}) - (1\text{f}) \right\}, \tag{6}$$

where $\widehat{\boldsymbol{\xi}}^\ell = (\widehat{\boldsymbol{a}}^\ell, \widehat{\boldsymbol{\lambda}}^\ell)$ *with* $\widehat{a}_{ij}^\ell = \max\{0, \bar{a}_{ij}^\ell - \theta/\gamma\}$ *and* $\widehat{\lambda}_{ij}^\ell = \max\{0, \bar{\lambda}_{ij}^\ell + \lfloor\theta\rfloor\}$.

Proof: The simplification of the inner supremum in DrFRAM (5) follows the monotonicity results in Corollary 1 and the definition of the weighted ℓ_∞ norm. □

Proposition 3 removes an obstacle in DrFRAM (5) by finding a closed-form worst-case representation for each scenario. Thus, we will mainly focus on addressing the second obstacle, which is the nonlinearity and non-convexity in the second-stage problem.

4. Linearizing the Second-stage Problem

As mentioned in the previous section, the difficulty of DrFRAM (6) resides in the second-stage problem (2) (i.e., the representation of the recourse function). The linearization technique is mainly based on the fact that the product of a binary variable $x \in \{0,1\}$ and a bounded continuous variable $y \in \{l, u\}$ can be linearized using the well-known McCormick inequalities (McCormick 1976), i.e.,

$$\{xy : x \in \{0,1\}, y \in [l,u]\} = \{z : y - u(1-x) \le z \le y - l(1-x), lx \le z \le ux, x \in \{0,1\}, y \in [l,u]\}.$$

Hence, to begin with, we first binarize the first-stage integer variables $\boldsymbol{n}$ as

$$n_{ik} = \sum_{r\in[R_k]} 2^{r-1} u_{ikr}, \forall i \in [I], \forall k \in [K], \tag{7a}$$

$$u_{ikr} \in \{0,1\}, \forall i \in [I], \forall k \in [K], \forall r \in [R_k], \tag{7b}$$

where we let $R_k = \lceil \log(\eta_k) \rceil$ for each $k \in [K]$ denote the largest possible bit when representing integer variable n_{ik}.

4.1. Linearizing Constraints (2b)

In this subsection, we focus on linearizing the constraints (2b) when $\boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}^\ell$ for each $\ell \in [N]$. First, to suppress the notation, we let $\boldsymbol{L}^\ell := \boldsymbol{L}(\widehat{\boldsymbol{\xi}}^\ell)$.

Before we linearize the constraints (2b), we observe that due to the monotonicity results in Part (i) of Proposition 2, constraints (2b) are equivalent to

$$L_{ij}^{\ell} \geq \min \left\{ \sum_{k\in[K]} n_{ik} \lfloor \delta_k c_k \rfloor, \lceil (1-\widehat{a}_{ij}^{\ell}) L_{i,j-1}^{\ell} \rceil + \lambda_{ij} \right\}, \forall i \in [I], \forall j \in [J_i]. \tag{8}$$

Since the ceiling function is lower semi-continuous, we can replace $\lceil (1-a_{ij}) L_{i,j-1}^{\ell} \rceil$ by its epigraph variable $\bar{L}_{i,j-1}^{\ell}$ such that

$$\bar{L}_{i,j-1}^{\ell} \geq (1-\widehat{a}_{ij}^{\ell}) L_{i,j-1}^{\ell}, \bar{L}_{i,j-1}^{\ell} \in \mathbb{Z}_+, \forall i \in [I], \forall j \in [J_i]. \tag{9}$$

In this way, constraints (8) can be further reformulated as

$$L_{ij}^{\ell} \geq \min \left\{ \sum_{k\in[K]} n_{ik} \lfloor \delta_k c_k \rfloor, \bar{L}_{i,j-1}^{\ell} + \widehat{\lambda}_{ij}^{\ell} \right\}, \forall i \in [I], \forall j \in [J_i],$$

Above, the piecewise minimum function can be linearized using binary variables $\boldsymbol{y}^{\ell}$, indicating whether each stop is fully occupied or not:

$$\begin{array}{l} L_{ij}^{\ell} \geq \sum_{k\in[K]} n_{ik} \lfloor \delta_k c_k \rfloor - M y_{ij}^{\ell}, L_{ij}^{\ell} \geq \bar{L}_{i,j-1}^{\ell} + \widehat{\lambda}_{ij}^{\ell} - M(1-y_{ij}^{\ell}), y_{ij}^{\ell} \in \{0,1\}, \forall i \in [I], \forall j \in [J_i], \\ L_{ij}^{\ell} \geq \sum_{k\in[K]} n_{ik} \lfloor \delta_k c_k \rfloor - (M-\widehat{\lambda}_{ij}^{\ell}) y_{ij}^{\ell}, L_{ij}^{\ell} \geq \bar{L}_{i,j-1}^{\ell} + \widehat{\lambda}_{ij}^{\ell} y_{ij}^{\ell}, y_{ij}^{\ell} \in \{0,1\}, \forall i \in [I], \forall j \in [J_i], \end{array} \tag{10}$$

where we choose $M = \max_{k\in[K]} \eta_k \lfloor \delta_k c_k \rfloor$ to be the maximum number of passengers that a route can carry at the same time. The equivalence of (8) and (10) is due to the fact that $\bar{L}_{i,j-1}^{\ell} \leq L_{i,j-1}^{\ell}$.

4.2. Linearizing the Second-stage Objective Function (2a)

For each $\ell \in [N]$, let us use E_1^{ℓ}, E_2^{ℓ} to denote the first and second parts of the objective function (2a). Then, according to constraints (9), equivalently, the second-stage objective function can be rewritten as

$$Q(\boldsymbol{n}, \boldsymbol{x}, \widehat{\boldsymbol{\xi}}^{\ell}, \widehat{\boldsymbol{L}}^{\ell}) = E_1^{\ell} + \omega E_2^{\ell}$$

where

$$E_1^{\ell} \left(\sum_{k\in[K]} n_{ik} \lfloor \delta_k c_k \rfloor \right) \geq L_{ij}^{\ell}, \forall i \in [I], \forall j \in [J_i], E_1^{\ell} \in [0,1], \tag{11a}$$

$$E_2^{\ell} \geq (\widehat{\lambda}_{ij}^{\ell})^{-1} \left(\bar{L}_{i,j-1}^{\ell} + \widehat{\lambda}_{ij}^{\ell} - \sum_{k\in[K]} n_{ik} \lfloor \delta_k c_k \rfloor \right), \forall i \in [I], \forall j \in [J_i], E_2^{\ell} \in [0,1]. \tag{11b}$$

Now it remains to linearize the bilinear terms in (11a). According to (7), we can represent variables $\boldsymbol{n}$ using binary variables $\boldsymbol{u}$. Therefore, constraints (11a) are equivalent to

$$\sum_{k\in[K]} \sum_{r\in[R_k]} 2^{r-1} \lfloor \delta_k c_k \rfloor E_1^{\ell} u_{ikr} \geq L_{ij}^{\ell}, \forall i \in [I], \forall j \in [J_i], E_1^{\ell} \in [0,1],$$

Introducing McCormick inequalities to linearize the bilinear terms, we have

$$\sum_{k\in[K]}\sum_{r\in[R_k]} 2^{r-1}\lfloor \delta_k c_k \rfloor w^\ell_{ikr} \geq L^\ell_{ij}, \forall i\in[I], \forall j\in[J_i], E^\ell_1 \in [0,1], \tag{11c}$$

$$w^\ell_{ikr}\geq 0, w^\ell_{ikr}\geq E^\ell_1 + u_{ikr} - 1, w^\ell_{ikr}\leq E^\ell_1, w^\ell_{ikr}\leq u_{ikr}, \forall k\in[K], \forall i\in[I], \forall r\in[R]. \tag{11d}$$

4.3. An Exact Mixed-Integer Linear Programming (MILP) Reformulation for DrFRAM (6)

Let us put all the linearized pieces together, and we are ready to present an exact MILP reformulation for DrFRAM (6). In particular, DrFRAM (6) is equivalent to the following MILP:

$$\text{(DrFRAM)} \qquad v^* = \min_{\boldsymbol{n},\boldsymbol{x},\boldsymbol{u}} \left\{ \frac{1}{N}\sum_{\ell\in[N]} \mathcal{Q}\left(\boldsymbol{n},\boldsymbol{x},\boldsymbol{u},\widehat{\boldsymbol{\xi}}^\ell\right) : (1\text{b})-(1\text{f}),(7\text{a}),(7\text{b}) \right\}, \tag{12a}$$

where for simplicity, we slightly abuse the notation by redefining $\mathcal{Q}(\boldsymbol{n},\boldsymbol{x},\boldsymbol{u},\widehat{\boldsymbol{\xi}}^\ell)$ as

$$\mathcal{Q}(\boldsymbol{n},\boldsymbol{x},\boldsymbol{u},\widehat{\boldsymbol{\xi}}^\ell) = \min_{\boldsymbol{L}^\ell,\bar{\boldsymbol{L}}^\ell,\boldsymbol{w}^\ell,\boldsymbol{y}^\ell,\boldsymbol{E}^\ell} \left\{ E^\ell_1 + \omega E^\ell_2 : (2\text{c}),(9),(10),(11\text{b})-(11\text{d}) \right\}. \tag{12b}$$

Note that we can encode the entire DrFRAM (12) into the off-the-shelf solvers such as Gurobi, CPLEX, MOSEK. However, our numerical study shows that, albeit effective, model (12) has difficulty solving large-scale instances, remaining large optimality gaps within an hour. Thus, we will develop valid inequalities and strong formulations based on the knapsack polytope and the disjunctive programming (Balas 1979) in the next section.

Finally, we remark that one can adopt the scenario decomposition method proposed by Ahmed (2013) to solve the DrFRAM (12) in a decomposed way. The general idea is to completely decomposition MILP (12) into N subproblems, i.e., for each $\ell\in[N]$, we solve

$$v^\ell := \min_{\boldsymbol{n},\boldsymbol{x},\boldsymbol{u}} \left\{ \mathcal{Q}\left(\boldsymbol{n},\boldsymbol{x},\boldsymbol{u},\widehat{\boldsymbol{\xi}}^\ell\right) : (1\text{b})-(1\text{f}),(7\text{a}),(7\text{b}) \right\}. \tag{13}$$

Subproblem (13) can be accelerated with the valid inequalities and stronger formulations developed in the next section. Note that the average of their objective functions provides a lower bound of DrFRAM, and the first-stage decision obtained in each subproblem is also feasible to the original MILP (12). Hence, we can evaluate these decisions and choose the best one as an upper bound. Next, we cut off the obtained decisions from their corresponding subproblems using the no-good cut (Ahmed 2013), i.e., for a scenario ℓ, given a first-stage binary solution $(\bar{\boldsymbol{x}}^\ell,\bar{\boldsymbol{u}}^\ell)$ (we do not need to include $\bar{\boldsymbol{n}}^\ell$ since it can be represented by $\bar{\boldsymbol{u}}^\ell$), we add the following no-good cut

$$\sum_{i\in[I],k\in[K]} \left((1-\bar{x}^\ell_{ik})x_{ik} + \sum_{r\in[R_k]} (1-\bar{u}^\ell_{ikr})u_{ikr} \right) \geq 1$$

Algorithm 1: Scenario Decomposition for Solving DrFRAM

Initialization: Set $\tau = 0, \epsilon > 0$, $\mathcal{X}^\ell = \{(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}) : (1\text{b}) - (1\text{f}), (7\text{a}), (7\text{b})\}$ for all $\ell \in [N]$, $LB = -\infty$, $UB = \infty$;

while $UB - LB > \epsilon$ **do**

  for $\ell \in [N]$ **do**

    $(\bar{\boldsymbol{n}}^\ell, \bar{\boldsymbol{x}}^\ell, \bar{\boldsymbol{u}}^\ell) \in \arg\min\{\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}, \widehat{\boldsymbol{\xi}}^\ell) : (\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}) \in \mathcal{X}^\ell\}$;

  $LB = \max\{LB, \frac{1}{N}\sum_{\ell \in [N]} \mathcal{Q}(\bar{\boldsymbol{n}}^\ell, \bar{\boldsymbol{x}}^\ell, \bar{\boldsymbol{u}}^\ell, \widehat{\boldsymbol{\xi}}^\ell)\}$;

  $UB = \min\{UB, \min_{t \in [N]} \frac{1}{N}\sum_{\ell \in [N]} \mathcal{Q}(\bar{\boldsymbol{n}}^t, \bar{\boldsymbol{x}}^t, \bar{\boldsymbol{u}}^t, \widehat{\boldsymbol{\xi}}^\ell)\}$;

  $\mathcal{X}^\ell \to \mathcal{X}^\ell \setminus \{(\bar{\boldsymbol{n}}^\ell, \bar{\boldsymbol{x}}^\ell, \bar{\boldsymbol{u}}^\ell)\}$ for each $\ell \in [N]$;

  $\tau \leftarrow \tau + 1$

into the ℓth subproblem; and repeat the same procedure. We terminate the solution procedure when invoking a stopping criterion. The detailed implementation can be found in Algorithm 1.

We can further extend the scenario decomposition by bundling the scenarios (see, e.g., **?**). Instead of solving N subproblems for each scenario $\ell \in [N]$, we group similar scenarios into disjoint subsets $G_\nu \subset [N], \nu \in [\mathcal{N}]$ using K-means clustering algorithm, and the subsets consist of a partition of $[N]$. We follow the same procedure that completely decomposition MILP (12) into $\mathcal{N}$ subproblems, i.e., for each $\nu \in [\mathcal{N}]$, we solve

$$v^\nu := \min_{\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}} \left\{ \frac{1}{|G_\nu|} \sum_{\ell \in G_\nu} \mathcal{Q}\left(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}, \widehat{\boldsymbol{\xi}}^\ell\right) : (1\text{b}) - (1\text{f}), (7\text{a}), (7\text{b}) \right\}. \tag{14}$$

The detailed implementation can be found in Algorithm 2.

Algorithm 2: Scenario Decomposition with Grouping for Solving DrFRAM

Initialization: Set $\tau = 0, \epsilon > 0$, integer $\mathcal{N}$ (divisible of N), cluster G_ν, and set $\mathcal{X}^\nu = \bigcap_{\ell \in G_\nu} \{(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}) : (1\text{b}) - (1\text{f}), (7\text{a}), (7\text{b})\}$ for all $\nu \in [\mathcal{N}]$, $LB = -\infty$, $UB = \infty$;

while $UB - LB > \epsilon$ **do**

  for $\nu \in [\mathcal{N}]$ **do**

    $(\bar{\boldsymbol{n}}^\nu, \bar{\boldsymbol{x}}^\nu, \bar{\boldsymbol{u}}^\nu) \in \arg\min\left\{v^\nu = \frac{1}{|G_\nu|}\sum_{\ell \in G_\nu} \mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}, \widehat{\boldsymbol{\xi}}^\ell) : (\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}) \in \mathcal{X}^\nu\right\}$;

  $LB = \max\{LB, \frac{1}{\mathcal{N}}\sum_{\nu \in [\mathcal{N}]} v^\nu\}$;

  $UB = \min\{UB, \min_{t \in [\mathcal{N}]} \frac{1}{N}\sum_{\ell \in [N]} \mathcal{Q}(\bar{\boldsymbol{n}}^t, \bar{\boldsymbol{x}}^t, \bar{\boldsymbol{u}}^t, \widehat{\boldsymbol{\xi}}^\ell)\}$;

  $\mathcal{X}^\nu \to \mathcal{X}^\nu \setminus \{(\bar{\boldsymbol{n}}^\nu, \bar{\boldsymbol{x}}^\nu, \bar{\boldsymbol{u}}^\nu)\}$ for each $\ell \in [\mathcal{N}]$;

  $\tau \leftarrow \tau + 1$

Note that to accelerate the algorithm, rather than adding the no-good cut to each subproblem, we can add an objective cut to force the objective function to increase by at least a positive step, that is, for each $\ell \in [N]$, we add the following inequality

$$E_1^\ell + \omega E_2^\ell \geq \mathcal{Q}\left(\bar{\boldsymbol{n}}^\ell, \bar{\boldsymbol{x}}^\ell, \bar{\boldsymbol{u}}^\ell\right) + \widehat{\epsilon}^\ell \tag{15}$$

into the ℓth subproblem, where $(\bar{\boldsymbol{n}}^\ell, \bar{\boldsymbol{x}}^\ell, \bar{\boldsymbol{u}}^\ell)$ denotes an optimal solution of ℓth subproblem and $\widehat{\epsilon}^\ell > 0$. Instead of solving the problem to optimal, the accelerated scenario decomposition can obtain a suboptimal solution. The detailed implementation can be found in Algorithm 3.

Algorithm 3: Accelerated Scenario Decomposition for Solving DrFRAM

Initialization: Set $\tau = 0, \epsilon > 0$, $\mathcal{X}^\ell = \{(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}) : (1b) - (1f), (7a), (7b)\}$ for all $\ell \in [N]$, $\{\widehat{\epsilon}^\ell\}_{\ell \in [N]}$, $LB = -\infty$, $UB = \infty$;
while $UB - LB > \epsilon$ **do**
  for $\ell \in [N]$ **do**
    $(\bar{\boldsymbol{n}}^\ell, \bar{\boldsymbol{x}}^\ell, \bar{\boldsymbol{u}}^\ell) \in \arg\min\{\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}, \widehat{\boldsymbol{\xi}}^\ell) : (15), (\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}) \in \mathcal{X}^\ell\}$;
  $LB = \max\{LB, \frac{1}{N}\sum_{\ell \in [N]} \mathcal{Q}(\bar{\boldsymbol{n}}^\ell, \bar{\boldsymbol{x}}^\ell, \bar{\boldsymbol{u}}^\ell, \widehat{\boldsymbol{\xi}}^\ell)\}$;
  $UB = \min\{UB, \min_{t \in [N]} \frac{1}{N}\sum_{\ell \in [N]} \mathcal{Q}(\bar{\boldsymbol{n}}^t, \bar{\boldsymbol{x}}^t, \bar{\boldsymbol{u}}^t, \widehat{\boldsymbol{\xi}}^\ell)\}$;
  $\tau \leftarrow \tau + 1$

Finally, we conclude this section by showing that if the data from distributionally robust counterparts (i.e., $\{\widehat{\boldsymbol{\xi}}^\ell\}_{\ell \in [N]}$) have a small variability, then the naive scenario decomposition lower bound (i.e., $1/N \sum_{\ell \in [N]} v^\ell$ will not be faraway from the true optimal v^*.

Proposition 4 *Suppose that the weight $\omega = 0$ (i.e., the passenger abandon rate for each bus stop is negligible), and there exists a matrix $\boldsymbol{\mu} = (\boldsymbol{\mu}^a, \boldsymbol{\mu}^\lambda) \in \Xi$ and a positive integer $\widehat{\alpha} \in \mathbb{Z}_{++}$ such that*

$$\widehat{a}_{ij}^\ell = \mu_{ij}^a, \mu_{ij}^\lambda \leq |\widehat{\lambda}_{ij}^\ell| \leq \widehat{\alpha}\mu_{ij}^\lambda, \forall i \in [I], \forall j \in [J_i].$$

Then the following approximation ratio holds for the naive scenario decomposition lower bound (without any no good cut or objective cut)

$$\frac{1}{\widehat{\alpha}} v^* \leq \frac{1}{N} \sum_{\ell \in [N]} v^\ell \leq v^*.$$

Proposition 4 shows that not very surprisingly, if the alighting rate for each bus stop is deterministic, and the passenger demand does not vary too much from scenario to scenario, then the naive scenario decomposition lower bound tends to be very close to the true optimality. This is consistent

with what we have found in the numerical study section. However, in general, Algorithm 1 suffers from slow convergence when being employed to find a global optimal solution. Therefore, in the next section, we will derive stronger formulations and valid inequalities to further strengthen MILP (12).

5. Stronger Formulations and Valid Inequalities

In this section, we strengthen the MILP model (12) by developing different families of valid inequalities. Our results and strong formulations are based on studying the mixed-integer substructures of model (12). Specifically, we consider convexifying the following two substructures to derive stronger formulations. First of all, for each $i \in [I]$, we consider the substructure of variables $(\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::})$ by projecting out variables $\boldsymbol{n}$. That is, for each $i \in [I]$, let us define

$$X_i^1 = \{(\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}) : \exists \boldsymbol{n}_{i:}, (1b) - (1c), (7a), (7b)\}. \tag{16a}$$

The second substructure is defined for each sample $\ell \in [N]$ and each $i \in [I]$ and $j \in [J_i]$ as

$$X_{ij\ell}^2 = \left\{(\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}, L_{ij}^\ell) \in X_i^1 \times \mathbb{R}_+ : (11c), (11d)\right\}. \tag{16b}$$

5.1. Convexification of Set X_i^1

We will first convexify X_i^1. According to (16a), set X_i^1 is equivalent to

$$X_i^1 = \left(\bigvee_{k \in [K]} \left(X_{ik}^1 \wedge \{x_{ik} = 1\}\right) \right) \bigwedge \left\{ \boldsymbol{x}_{i:} \in \{0,1\}^K : \sum_{k \in [K]} x_{ik} = 1 \right\}$$

where

$$X_{ik}^1 := \left\{ (\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}) : \sum_{r \in [R_k]} 2^{r-1} u_{ikr} \le \eta_k x_{ik}, u_{ikr} \in \{0,1\}, \forall r \in [R_k] \right\}.$$

Suppose that $\eta_k = 2^{i_{1k}-1} + \ldots + 2^{i_{sk}-1} + 2^{R_k - 1}$ and set $\mathcal{I} := \{i_{1k}, \ldots, i_{sk}, R_k\}$. Since $\{2^{r-1}\}_{r \in [R_k]}$ are superincreasing, according to proposition 3.4 in Gupte et al. (2013), the convex hull of $X_{ik}^1 \wedge \{x_{ik} = 1\}$ can be described as

$$\operatorname{conv}\left(X_{ik}^1 \wedge \{x_{ik} = 1\}\right) = \left\{ (\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}) : \begin{array}{r} \boldsymbol{x}_{i:} \in [0,1]^K, x_{ik} = 1, \boldsymbol{u}_{ik:} \in [0,1]^{R_k}, \\ u_{ikr} + \sum_{\tau \in \mathcal{I}_r} u_{ik\tau} \le |\mathcal{I}_r|, \forall r \in [R_k] \setminus \mathcal{I} \end{array} \right\}$$

where $\mathcal{I}_r := \{s \in \mathcal{I} : s > r\}$.

Our convex hull description of set X_i^1 relies on the well-known disjunctive programming (Balas 1979).

Lemma 1 *Suppose $\bar{s}$ polyhedra $\bar{X}^i = \{\bar{\boldsymbol{y}} \in \mathbb{R}^{\bar{n}} : \boldsymbol{A}^i \bar{\boldsymbol{y}} \leq \boldsymbol{b}^i\}$ with $\boldsymbol{A}^i \in \mathbb{R}^{\bar{m}_i \times \bar{n}}$ and $\boldsymbol{b} \in \mathbb{R}^{\bar{m}_i}$ for each $i \in [\bar{s}]$ share the same recession cone. Then the following result holds*

$$\begin{aligned}&\operatorname{conv}\left\{\left(\bigvee_{s\in[\bar{s}]}\left(\bar{X}^i \wedge \{\lambda_i = 1\}\right)\right) \bigwedge \left\{\boldsymbol{\lambda} \in \{0,1\}^{\bar{s}} : \sum_{i\in[\bar{s}]} \lambda_i = 1\right\}\right\}\\ &= \left\{(\bar{\boldsymbol{y}}, \boldsymbol{\lambda}) \in \mathbb{R}^{\bar{n}} \times [0,1]^{\bar{s}} : \boldsymbol{A}^i \bar{\boldsymbol{y}}^i \leq \boldsymbol{b}^i \lambda_i, \forall i \in [\bar{s}], \sum_{i\in[\bar{s}]} \bar{\boldsymbol{y}}^i = \bar{\boldsymbol{y}}\right\}.\end{aligned}$$

According to Lemma 1 and the fact that the convex hull of the union of sets is equal to the convex hull of the union of the convex hulls of sets, we arrive at the complete description of the convex hull of set X_i^1. This result is summarized below.

Proposition 5 *For each $i \in [I]$, the convex hull of the set X_i^1 is equal to*

$$\operatorname{conv}\left(X_i^1\right) = \left\{(\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}) : \begin{array}{c} \boldsymbol{x}_{i:} \in [0,1]^K, \sum_{k\in[K]} x_{ik} = 1, \\ u_{ikr} + \sum_{\tau\in\mathcal{I}_r} u_{ik\tau} \leq |\mathcal{I}_r| x_{ik}, \forall r \in [R_k] \setminus \mathcal{I}, \forall k \in [K], \\ 0 \leq u_{ikr} \leq x_{ik}, \forall r \in [R_k], \forall k \in [K] \end{array}\right\}. \tag{17}$$

5.2. Convexification of Set $X_{ij\ell}^2$

Similarly, to convexify set $X_{ij\ell}^2$, we first rewrite it as a disjunction

$$X_{ij\ell}^2 = \left(\bigvee_{k\in[K]}\left(X_{ij\ell k}^2 \wedge \{x_{ik} = 1\}\right)\right) \bigwedge \left\{\boldsymbol{x}_{i:} \in \{0,1\}^K : \sum_{k\in[K]} x_{ik} = 1\right\}, \tag{18}$$

where

$$X_{ij\ell k}^2 \wedge \{x_{ik} = 1\} := \left\{(\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}, L_{ij}^\ell, \boldsymbol{w}_{i::}^\ell, E_1^\ell) : \begin{array}{c} (\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}) \in X_{ik}^1, \sum_{r\in[R_k]} 2^{r-1} \lfloor \delta_k c_k \rfloor w_{ikr}^\ell \geq L_{ij}^\ell, \\ w_{ikr}^\ell = u_{ikr} E_1^\ell, L_{ij}^\ell \in \mathbb{R}_+, E_1^\ell \in [0,1], x_{ik} = 1 \end{array}\right\}.$$

We first observe that the convex hull of set $X_{ij\ell k}^2 \wedge \{x_{ik} = 1\}$ is equal to the intersection of the constraint $\sum_{r\in[R_k]} 2^{r-1} \lfloor \delta_k c_k \rfloor w_{ikr}^\ell \geq L_{ij}^\ell$ and the convex hull of the feasible region subject to the remaining constraints.

Lemma 2 *The following characterization holds for $\operatorname{conv}(X_{ij\ell k}^2 \wedge \{x_{ik} = 1\})$, i.e.,*

$$\operatorname{conv}(X_{ij\ell k}^2 \wedge \{x_{ik} = 1\}) = \operatorname{conv}(\bar{X}_{ij\ell k}^2) \bigcap \left\{(\boldsymbol{w}_{i::}^\ell, L_{ij}^\ell) : \sum_{r\in[R_k]} 2^{r-1} \lfloor \delta_k c_k \rfloor w_{ikr}^\ell \geq L_{ij}^\ell\right\},$$

where

$$\bar{X}_{ij\ell k}^2 = \left\{(\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}, L_{ij}^\ell, \boldsymbol{w}_{i::}^\ell, E_1^\ell) : (\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}) \in X_{ik}^1, w_{ikr}^\ell = u_{ikr} E_1^\ell, L_{ij}^\ell \in \mathbb{R}_+, E_1^\ell \in [0,1], x_{ik} = 1\right\}.$$

It turns out that according to Lemma 2 and proposition 3.1. in Gupte et al. (2013), sets $\text{conv}(\bar{X}^2_{ij\ell k})$ and $\text{conv}(X^2_{ij\ell k} \wedge \{x_{ik} = 1\})$ can be completely described using the result in Proposition 5 and the fact that variables $\boldsymbol{x}_{i:}$ are independent of others and constraint system with respect to $\boldsymbol{x}_{i:}$ are integral. These results are summarized below.

Proposition 6 *For each* $i \in [I], j \in [J_i], k \in [K], \ell \in [N]$*, the sets* $\text{conv}(\bar{X}^2_{ij\ell k})$ *and* $\text{conv}(X^2_{ij\ell k} \wedge \{x_{ik} = 1\})$ *admit the following complete descriptions:*

- $$\text{conv}(\bar{X}^2_{ij\ell k}) = \left\{ (\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}, L^\ell_{ij}, \boldsymbol{w}^\ell_{i::}, E^\ell_1) : \begin{array}{l} \boldsymbol{x}_{i:} \in [0,1]^K, x_{ik} = 1, E^\ell_1 \in [0,1], \\ w_{ikr} + \sum_{\tau \in \mathcal{I}_r} w_{ik\tau} \le |\mathcal{I}_r| E^\ell_1, \forall r \in [R_k] \setminus \mathcal{I}, \\ u_{ikr} + \sum_{\tau \in \mathcal{I}_r} u_{ik\tau} - w_{ikr} - \sum_{\tau \in \mathcal{I}_r} w_{ik\tau} \le \\ |\mathcal{I}_r|(1 - E^\ell_1), \forall r \in [R_k] \setminus \mathcal{I}, \\ 0 \le w_{ikr} \le E^\ell_1, \forall r \in [R_k], \\ 0 \le u_{ikr} - w_{ikr} \le (1 - E^\ell_1), \forall r \in [R_k] \end{array} \right\}; \tag{19}$$

  and

- $$\text{conv}(X^2_{ij\ell k} \wedge \{x_{ik} = 1\}) = \left\{ (\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}, L^\ell_{ij}, \boldsymbol{w}^\ell_{i::}, E^\ell_1) : \begin{array}{l} \boldsymbol{x}_{i:} \in [0,1]^K, x_{ik} = 1, E^\ell_1 \in [0,1], \\ w_{ikr} + \sum_{\tau \in \mathcal{I}_r} w_{ik\tau} \le |\mathcal{I}_r| E^\ell_1, \forall r \in [R_k] \setminus \mathcal{I}, \\ u_{ikr} + \sum_{\tau \in \mathcal{I}_r} u_{ik\tau} - w_{ikr} - \sum_{\tau \in \mathcal{I}_r} w_{ik\tau} \le \\ |\mathcal{I}_r|(1 - E^\ell_1), \forall r \in [R_k] \setminus \mathcal{I}, \\ 0 \le w_{ikr} \le E^\ell_1, \forall r \in [R_k], \\ 0 \le u_{ikr} - w_{ikr} \le (1 - E^\ell_1), \forall r \in [R_k], \\ \sum_{r \in [R_k]} 2^{r-1} \lfloor \delta_k c_k \rfloor w^\ell_{ikr} \ge L^\ell_{ij} \end{array} \right\}. \tag{20}$$

Following Lemma 1 and the fact that the convex hull of a union of sets is equal to the convex hull of the union of convex hulls of sets, we arrive at the following complete description of the convex hull of set $X^2_{ij\ell}$.

Proposition 7 *For each $i \in [I], j \in [J_i]$, and $\ell \in [N]$, the set $\mathrm{conv}(X^2_{ij\ell})$ admits the following complete description:*

$$\mathrm{conv}(X^2_{ij\ell}) = \left\{ (\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}, L^\ell_{ij}, \boldsymbol{w}^\ell_{i::}, E^\ell_1) : \begin{array}{l} \boldsymbol{x}_{i:} \in [0,1]^K, \sum_{k\in[K]} x_{ik} = 1, \sum_{k\in[K]} \widehat{E}^\ell_{1ik} = E^\ell_1, \widehat{E}^\ell_{1ik} \le x_{ik}, \\ w_{ikr} + \sum_{\tau\in\mathcal{I}_r} w_{ik\tau} \le |\mathcal{I}_r| \widehat{E}^\ell_{1ik}, \forall r \in [R_k] \setminus \mathcal{I}, \forall k \in [K], \\ u_{ikr} + \sum_{\tau\in\mathcal{I}_r} u_{ik\tau} - w_{ikr} - \sum_{\tau\in\mathcal{I}_r} w_{ik\tau} \le \\ |\mathcal{I}_r|(1 - \widehat{E}^\ell_{1ik}), \forall r \in [R_k] \setminus \mathcal{I}, \\ 0 \le w_{ikr} \le \widehat{E}^\ell_{1ik}, \forall r \in [R_k], \forall k \in [K], \\ 0 \le u_{ikr} - w_{ikr} \le (1 - \widehat{E}^\ell_{1ik}), \forall r \in [R_k], \forall k \in [K], \\ \sum_{r\in[R_k]} 2^{r-1} \lfloor \delta_k c_k \rfloor w^\ell_{ikr} \ge \widehat{L}^\ell_{ijk}, \forall k \in [K], \\ \sum_{k\in[K]} \widehat{L}^\ell_{ijk} = L^\ell_{ij}, 0 \le \widehat{E}^\ell_{1ik} \le E^\ell_1 \le 1, \widehat{L}^\ell_{ijk} \ge 0, \forall k \in [K] \end{array} \right\}. \tag{21}$$

5.3. Valid Inequalities

In this subsection, we develop valid inequalities to further strengthen the MILP model (12).

One-bus-per-route Inequalities: First, at-least-one-bus-per-route constraints (1e) together with the binarization constraints (7a) imply that at least one of the corresponding binary variables must be nonzero, i.e., the following valid inequalities must hold:

$$\sum_{r\in[R_k]} u_{ikr} \ge x_{ik}, \forall i \in [I], \forall k \in [K]. \tag{22}$$

Under-capacity Inequalities: Let us define $\underline{M} = \min_{k\in[K]} \lfloor \delta_k c_k \rfloor$ as the minimum route capacity at any route. Suppose each route has no capacity restriction, i.e., any passenger is supposed to get on board, then we can compute the ideal number of passengers remaining on the bus using the following formula:

$$\iota^\ell_{ij} = \lceil (1 - \widehat{a}^\ell_{ij}) \iota^\ell_{i,j-1} \rceil + \widehat{\lambda}^\ell_{ij}, \forall i \in [I], \forall j \in [J_i], \forall \ell \in [N], \tag{23}$$

where we let $\iota^\ell_{i0} = 0$ for all $i \in [I]$ and $\ell \in [N]$. Now let us define $j^\ell_{i*} = \min\{J_i, \min_{j\in[J_i]}\{\iota^\ell_{i0} > \underline{M}\}$ to be the first bus stop such that the number of passengers planning to be on board is greater than the minimum capacity for each $i \in [I]$ and $\ell \in [N]$. Then, clearly, for any bus stop $j \in [j^\ell_{i*} - 1]$, it is under-capacity, i.e., we must have

$$y^\ell_{ij} = 1, \forall i \in [I], \forall j \in [j^\ell_{i*} - 1], \forall \ell \in [N]. \tag{24}$$

Utilization Rate when Being Fully Occupied: If one bus stop is fully occupied, the bus utilization rate must be one. Thus, we have the following valid inequalities:

$$1 - y^\ell_{ij} \le E^\ell_1, \forall i \in [I], \forall j \in [J_i], \forall \ell \in [N]. \tag{25}$$

Tightening Big-M Coefficients: Note that in (10), we must have $L_{ij}^{\ell} \geq \widehat{\lambda}_{ij}^{\ell}$ if $\widehat{y}_{ij}^{\ell} = 1$ and $L_{ij}^{\ell} \geq \bar{L}_{i,j-1}^{\ell}$. Hence, in the first part of (10), we can reduce M by $M - \widehat{\lambda}_{ij}^{\ell}$, and in the second part of (10), $\widehat{\lambda}_{ij}^{\ell} - M(1 - y_{ij}^{\ell})$ can be tightened by $\widehat{\lambda}_{ij}^{\ell} y_{ij}^{\ell}$. Therefore, we arrive at the following stronger inequalities

$$L_{ij}^{\ell} \geq \sum_{k\in[K]} n_{ik}\lfloor \delta_k c_k \rfloor - (M - \widehat{\lambda}_{ij}^{\ell}) y_{ij}^{\ell}, L_{ij}^{\ell} \geq \bar{L}_{i,j-1}^{\ell} + \widehat{\lambda}_{ij}^{\ell} y_{ij}^{\ell}, \forall i \in [I], \forall j \in [J_i], \forall \ell \in [N]. \tag{26}$$

Passengers Being Unserved when a Bus is Fully Occupied: We note that the passenger abandon rate becomes positive only when the bus is fully occupied at some moments, i.e., a passenger will be allowed on board unless the bus is full. Thus, we must have

$$E_2^{\ell} \leq E_1^{\ell}, \forall \ell \in [N]. \tag{27}$$

Lower Bounding the Bus Utilization Rate: The disjunctive programming results in Section 5.2 implies that $\widehat{E}_{1ik}^{\ell} \geq \widehat{L}_{ijk}^{\ell}/(n_{ik}\lfloor \delta_k c_k \rfloor)$. Since $n_{ik} \leq \widehat{M}_k := \min\{\eta_k, \sum_{\kappa\in[K]} \eta_\kappa - I + 1\}$, we must have

$$\widehat{E}_{1ik}^{\ell} \geq \frac{\widehat{L}_{ijk}^{\ell}}{\widehat{M}_k \lfloor \delta_k c_k \rfloor}, \forall i \in [I], \forall j \in [J_i], \forall \ell \in [N]. \tag{28}$$

6. Approximation Algorithms Based on No-One-Left Policy

We observe that the intricacy of the MILP model (12) comes from the linearization of the first part of the objective function using binary variables $\boldsymbol{u}$ and the linearization of constraints (2b) using auxiliary variables $\boldsymbol{y}$. To avoid both, we propose an approximation scheme using the notion of the so-called No-one-left policy. Recall that in the previous section, we define $\boldsymbol{\iota}$ to be the ideal number of passengers at each bus stop when there is no capacity restriction. Our No-one-left policy follows the same concept.

Thus, we propose to solve the following simplified model as

$$(\textbf{No-one-left}) \qquad \min_{\boldsymbol{n},\boldsymbol{x}} \left\{ \frac{1}{N} \sum_{\ell\in[N]} \overline{\mathcal{Q}}\left(\boldsymbol{n}, \boldsymbol{x}, \widehat{\boldsymbol{\xi}}^{\ell}\right) : (1b) - (1f) \right\}, \tag{29a}$$

where we define $\overline{\mathcal{Q}}(\boldsymbol{n}, \boldsymbol{x}, \widehat{\boldsymbol{\xi}}^{\ell}) = \max_{i\in[I], j\in[J_i]} \iota_{ij}^{\ell}/(\sum_{k\in[K]} n_{ik}\lfloor \delta_k c_k \rfloor)$. Note that $\overline{\mathcal{Q}}(\boldsymbol{n}, \boldsymbol{x}, \widehat{\boldsymbol{\xi}}^{\ell})$ admits a second-order conic representation as

$$\overline{\mathcal{Q}}(\boldsymbol{n}, \boldsymbol{x}, \widehat{\boldsymbol{\xi}}^{\ell}) = \min_{\overline{E}^{\ell}\in\mathbb{R}_+} \left\{ \overline{E}^{\ell} : \left\| \left(2\sqrt{\max_{j\in[J_i]} \iota_{ij}^{\ell}}, \sum_{k\in[K]} n_{ik}\lfloor \delta_k c_k \rfloor - \overline{E}^{\ell} \right) \right\|_2 \leq \sum_{k\in[K]} n_{ik}\lfloor \delta_k c_k \rfloor + \overline{E}^{\ell}, \forall i \in [I] \right\}. \tag{29b}$$

After solving the No-one-left model (29), we can evaluate its objective value v^N by plugging in its optimal first-stage decision $(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}})$ into DrFRAM (6), i.e.,

$$v^N = \frac{1}{N} \sum_{\ell \in [N]} \mathcal{Q}\left(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}}, \widehat{\boldsymbol{\xi}}^\ell\right).$$

The following result shows the approximation bound of the No-one-left policy.

Proposition 8 *Suppose that* $(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}}, \overline{\boldsymbol{E}})$ *and* $(\boldsymbol{n}^*, \boldsymbol{x}^*, \boldsymbol{E}^*)$ *are optimal solutions to the No-one-left model* (29) *and DrFRAM* (12)*, respectively. Then we have*

$$v^* \le v^N \le v^* + \frac{\omega}{N} \|(\overline{\boldsymbol{E}} - \mathbf{e})_+\|_0 + \frac{\overline{M}/\underline{M} - 1}{N} \|\boldsymbol{E}_2^*\|_0,$$

where $\underline{M} := \min_{k \in [K]} \lfloor \delta_k c_k \rfloor$ *and* $\overline{M} := \max_{i \in [I], j \in [J_i], \ell \in [N]} \iota_{ij}^\ell$.

Proof: First of all, $v^* \le v^N$ is due to the feasibility of the first stage decision $(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}})$.

Next, we note that if $\overline{E}^\ell > 1$, then we have $1 \le \mathcal{Q}\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}}, \widehat{\boldsymbol{\xi}}^\ell) \le 1 + \omega$. Therefore, we must have

$$v^N - \frac{\omega}{N} \|(\overline{\boldsymbol{E}} - \mathbf{e})_+\|_0 \le \frac{1}{N} \sum_{\ell \in [N]} \overline{E}^\ell. \tag{30}$$

According to the optimality of $(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}}, \overline{\boldsymbol{E}})$ to No-one-left model (29), we have

$$\frac{1}{N} \sum_{\ell \in [N]} \overline{E}^\ell \le \frac{1}{N} \sum_{\ell \in [N]} \frac{\max_{i \in [I], j \in [J_i]} \iota_{ij}^\ell}{\sum_{k \in [K]} n_{ik}^* \lfloor \delta_k c_k \rfloor} \le v^* + \frac{\overline{M}/\underline{M} - 1}{N} \|\boldsymbol{E}_2^*\|_0. \tag{31}$$

Combining these two inequalities (30) and (31) together, we arrive at the conclusion. □

Proposition 8 shows that $(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}})$ is optimal to DrFRAM (6) if no passenger is being abandoned (i.e., $\boldsymbol{E}_2^* = \mathbf{0}$) in an optimal solution to the DrFRAM model and all the utilization rates are zero in the No-one-left model.

6.1. Enhancing No-one-left Policy

In this subsection, we propose to enhance the No-one-left policy by incorporating the abandon rate for each scenario that reaches a full utilization rate of some routes and resolving the problems again. Namely, suppose $(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}})$ denotes an optimal first-stage decision of the No-one-left model (29). Let us denote two subsets of scenarios, $\mathcal{N}_- = \{\ell \in [N] : \mathcal{Q}\left(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}}, \widehat{\boldsymbol{\xi}}^\ell\right) \le 1\}$ and $\mathcal{N}_+ = \{\ell \in [N] : \mathcal{Q}\left(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}}, \widehat{\boldsymbol{\xi}}^\ell\right) > 1\}$. Then the enhanced model is defined as follows:

$$(\textbf{EnModel}) \qquad \min_{\boldsymbol{n}, \boldsymbol{x}} \left\{ \frac{1}{N} \left[\sum_{\ell \in \mathcal{N}_-} \overline{\mathcal{Q}}\left(\boldsymbol{n}, \boldsymbol{x}, \widehat{\boldsymbol{\xi}}^\ell\right) + \sum_{\ell \in \mathcal{N}_+} \overline{\mathcal{Q}}^E\left(\boldsymbol{n}, \boldsymbol{x}, \widehat{\boldsymbol{\xi}}^\ell\right) \right] : (1\text{b}) - (1\text{f}) \right\}, \tag{32a}$$

where $\overline{\mathcal{Q}}(\boldsymbol{n}, \boldsymbol{x}, \widehat{\boldsymbol{\xi}}^\ell)$ is defined in (29) and

$$\overline{\mathcal{Q}}^E(\boldsymbol{n}, \boldsymbol{x}, \widehat{\boldsymbol{\xi}}^\ell) = \min_{\boldsymbol{L}^\ell, \bar{\boldsymbol{L}}^\ell, \boldsymbol{y}^\ell, E_2^\ell} \left\{ 1 + E_2^\ell : (9), (10), (11\text{b}), E_2^\ell \in [0, 1] \right\}. \tag{32b}$$

Once solving the EnModel (32), we can repeat the procedure until invoking the stopping criteria. This procedure is summarized in Algorithm 4.

Algorithm 4: Enhancing No-one-left Policy

Initialization: Solve the No-one-left model (29) with an optimal first-stage solution $(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}})$; set $t = 0, N_- = [N]$ and initialize $t_{\max}$;
while *Set* $\mathcal{N}_-$ *is changing and* $t < t_{\max}$ **do**
  Define $\mathcal{N}_- = \{\ell \in [N] : \mathcal{Q}(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}}, \widehat{\boldsymbol{\xi}}^\ell) \leq 1\}$ and $\mathcal{N}_+ = \{\ell \in [N] : \mathcal{Q}(\bar{\boldsymbol{n}}, \bar{\boldsymbol{x}}, \widehat{\boldsymbol{\xi}}^\ell) > 1\}$;
  Solve the EnModel (32);
  $t \leftarrow t + 1$.

7. Numerical Study

In this section, we present a set of numerical results to compare the strengths of different model formulations and test the effectiveness of distinct methods using both small and large random instances as well as real-world data provided by Blacksburg Transit. For the random instances, passenger arrival rates were generated from uniform distributions with the minimum value ranging from 2 to 20 and maximum value ranging from 22 to 40, and the proportion of passengers alighting from the bus was randomly generated from triangular distributions with the lower limit ranging from 0 to 0.2, the upper limit range from 0.6 to 1, and the mode ranging from 0.45 to 0.55. A time limit of 3600 seconds was set for solving each instance. All the instances were coded in Python 3.7.0 with calls to Gurobi 9.0.3 on a personal computer with a 1.9 GHz Intel Core i7 processer and 16G memory.

7.1. Results of Small Instances

In this section, we tested five small instances to compare the performances of MILP formulations as well as their continuous relaxations, where the number of bus routes I is 5 or 6 and the number of scenarios N is 5, and $K = 3$ types of buses with nominal capacities $c_1 = 60, c_2 = 80, c_3 = 120$. Each route has 10 to 40 stops. We generated 50 cases based on these instances using different parameter combinations (i.e., different values of $I, \boldsymbol{\eta}, \delta, \omega, \theta$). Particularly, cases 1-25, 26-35, 36-40, 41-45, 46-50 correspond to instances 1,2,3,4,5, respectively. For all the cases, we fixed the normalized parameter $\gamma = 100$.

The numerical results of exactly solving different MILP formulations are reported in Table 1, where $a-b$ in the "Case" column means instance $\#a$ and case $\#b$, "MILP.B" represents MILP (12), "OPT" represents the optimal objective value obtained, "MILP.VI" represents MILP (12) with (22)-(27), "MILP.CONV" represents MILP (12) with (17), (21), (26), and (28), "MILP.VI.CONV" represents MILP (12) with (17), (21), (22)-(28). Without a doubt, we see that all the formulations are able to find an optimal solution. MILP.VI improves MILP.B's solution time on average but introduces more nodes to explore. This is probably because the valid inequalities introduced may

force the solver to explore different branches before reaching an optimal solution. MILP.CONV requires fewer nodes to explore, but it usually takes more time to solve. This is possible because of additional variables introduced to describe the convex hulls. The running time of MILP.VI.CONV is in between that of MILP.VI and MILP.CONV, but it has the least number of nodes to explore. Hence, MILP.VI is the best among these methods.

We see that if the Wasserstein radius θ increases, the objective value increases, and solution time does not change too much, since any increase in θ may result in larger passenger arrival rates and smaller passenger alighting rates. As the pandemic factor δ or the total number of buses increases and other parameters stay the same, both objective value and solution time decrease, since larger δ or larger $\boldsymbol{\eta}$ implies larger capacities for some routes and thus more passengers to get onto the bus. Weight ω does not affect the objective value and solution time too much when the objective value is less than 1. When the objective value exceeds 1, a larger ω results in a larger objective value while the solution time stays the same. Since the problem size grows as the number of routes I increases, the solution time also increases.

Table 1 Results of Different MILP Formulations

Case	I	$\boldsymbol{\eta}$	δ	ω	θ	OPT	MILP.B		MILP.VI		MILP.CONV		MILP.VI.CONV	
							Time	Nodes	Time	Nodes	Time	Nodes	Time	Nodes
1 - 1					0	0.885	40	2624	27	3547	85	4221	44	2241
1 - 2					0.25	0.920	56	5123	30	3689	203	5284	77	2452
1 - 3	5	(4,6,2)	0.25	0.25	0.5	0.930	36	3781	14	10032	104	3308	99	3081
1 - 4					1	0.940	67	7065	40	7375	99	4823	79	2889
1 - 5					1.5	0.972	45	5870	28	6732	160	4912	112	2732
Average							49	4893	28	6275	130	4510	82	2679
1 - 6					0	0.885	24	3646	20	3841	69	2936	50	1611
1 - 7					0.25	0.920	30	2421	26	2588	64	2449	66	1905
1 - 8	5	(4,6,2)	0.25	0.5	0.5	0.930	21	2448	11	7222	63	2121	93	2667
1 - 9					1	0.940	36	2542	32	2984	87	2548	96	2173
1 - 10					1.5	0.982	49	5110	14	3126	86	2516	107	2348
Average							32	3233	21	3952	74	2514	82	2141
1 - 11					0	0.885	23	1955	39	4716	85	3691	55	1455
1 - 12					0.25	0.920	24	3101	21	2496	43	1402	57	1782
1 - 13	5	(4,6,2)	0.25	1	0.5	0.930	16	1682	6	3285	49	1786	76	2161
1 - 14					1	0.940	18	2186	46	4333	47	1363	106	2953
1 - 15					1.5	1.000	26	2598	18	4093	82	2668	118	2264
Average							21	2304	26	3785	61	2182	82	2123
1 - 16					0	0.572	20	3610	18	9968	37	6979	43	6797
1 - 17					0.25	0.596	8	5566	15	7568	92	2279	49	3650
1 - 18	5	(6,9,3)	0.25	0.5	0.5	0.600	24	2818	25	14153	126	14562	51	2210
1 - 19					1	0.609	32	4904	27	17539	116	2651	51	2930
1 - 20					1.5	0.632	36	4601	17	9853	92	2308	56	4023
Average							24	4300	20	11816	93	5756	50	3922
1 - 21					0	0.286	3	1856	2	1224	32	2649	17	1364
1 - 22					0.25	0.298	6	2965	3	3141	34	3712	16	2123
1 - 23	5	(6,9,3)	0.5	0.5	0.5	0.300	6	1514	6	2484	37	3381	50	4174
1 - 24					1	0.304	18	1809	2	1305	34	2573	27	1720
1 - 25					1.5	0.316	10	1705	2	2968	34	5101	76	21284
Average							9	1970	3	2224	34	3483	37	6133
2 - 26					0	0.389	53	7363	37	11959	164	28193	90	12605
2 - 27					0.25	0.404	111	12592	31	15186	440	8294	124	15668
2 - 28	6	(10,15,5)	0.25	0.5	0.5	0.404	75	6496	23	9392	334	31837	132	14800
2 - 29					1	0.408	47	5636	16	6465	338	33053	73	6650
2 - 30					1.5	0.427	37	4858	35	13029	415	28730	121	8573
Average							64	7389	28	11206	338	26021	108	11659
2 - 31					0	0.389	66	10349	18	7355	242	35119	67	9443
2 - 32					0.25	0.404	61	7711	27	11740	283	5371	95	12051
2 - 33	6	(10,15,5)	0.25	1	0.5	0.404	58	7054	24	9935	317	6177	76	7333
2 - 34					1	0.408	53	5675	23	8398	113	19688	62	8379
2 - 35					1.5	0.427	136	17187	28	13626	481	28128	198	11008
Average							75	9595	24	10211	287	18897	100	9643

Results of continuous relaxations of different MILP formulations are reported in Table 2, where we use "C.MILP.B," "C.MILP.VI," "C.MILP.CONV," "C.MILP.VI.CONV" to denote the continuous relaxations of MILP.B, MILP.VI, MILP.CONV, MILP.VI.CONV, respectively. We also define "Gap" as the percentage of the difference between the objective value and OPT divided by OPT. It is evident that both C.MILP.VI and C.MILP.CONV are better than C.MILP. This demonstrates the effectiveness of the proposed valid inequalities and the proposed convexification results. Clearly, the integration of valid inequalities and convexification results is the best, i.e., C.MILP.VI.CONV yields the smallest optimality gap, which is around 90% on average. Such results indicate that the MILP.VI.CONV model is the most effective in improving the root gap. We also notice that valid inequalities contribute more than convexification results in improving the optimality gap since C.MILP.VI is consistently better than C.MILP.CONV. Besides, C.MILP.CONV takes more time than C.MILP.VI. This also explains why MILP.VI performs better than MILP.CONV in Table 1. Regarding both solution time and the formulation strength, among the four different MILP formulations, MILP.VI tends to be the best. Since the continuous relaxation values of MILP.B are nearly 0, deriving stronger MILP formulations is of necessity.

Table 2 Results of Continuous Relaxations of Different MILP Models

Case	C.MILP.B			C.MILP.VI			C.MILP.CONV			C.MILP.VI.CONV		
	Obj.Val	Time	Gap(%)	Obj.Val	Time	Gap(%)	Obj.Val	Time	Gap(%)	Obj.Val	Time	Gap(%)
1 - 1	0.000	0.02	100.0	0.089	0.06	89.9	0.088	0.09	90.0	0.095	0.22	89.3
1 - 2	0.000	0.02	100.0	0.091	0.05	90.1	0.092	0.07	90.0	0.097	0.20	89.5
1 - 3	0.000	0.02	100.0	0.093	0.06	90.0	0.095	0.08	89.8	0.098	0.23	89.5
1 - 4	0.000	0.02	100.0	0.094	0.05	90.0	0.097	0.22	89.7	0.100	0.24	89.4
1 - 5	0.000	0.02	100.0	0.098	0.06	89.9	0.102	0.09	89.5	0.104	0.20	89.3
Average		0.02	100.0		0.06	90.0		0.11	89.8		0.22	89.4
1 - 6	0.000	0.02	100.0	0.089	0.08	89.9	0.091	0.17	89.7	0.095	0.23	89.3
1 - 7	0.000	0.02	100.0	0.091	0.06	90.1	0.096	0.09	89.5	0.097	0.21	89.5
1 - 8	0.000	0.02	100.0	0.093	0.06	90.0	0.099	0.11	89.4	0.098	0.22	89.5
1 - 9	0.000	0.02	100.0	0.094	0.08	90.0	0.101	0.10	89.3	0.100	0.21	89.4
1 - 10	0.000	0.02	100.0	0.098	0.06	90.0	0.107	0.08	89.1	0.104	0.21	89.4
Average		0.02	100.0		0.07	90.0		0.11	89.4		0.22	89.4
1 - 11	0.000	0.02	100.0	0.089	0.06	89.9	0.091	0.07	89.7	0.095	0.21	89.3
1 - 12	0.000	0.02	100.0	0.091	0.08	90.1	0.096	0.08	89.5	0.097	0.19	89.5
1 - 13	0.000	0.02	100.0	0.093	0.07	90.0	0.099	0.08	89.4	0.098	0.25	89.5
1 - 14	0.000	0.02	100.0	0.094	0.07	90.0	0.101	0.07	89.3	0.100	0.27	89.4
1 - 15	0.000	0.02	100.0	0.098	0.06	90.2	0.108	0.08	89.2	0.104	0.22	89.6
Average		0.02	100.0		0.07	90.0		0.08	89.4		0.23	89.5
1 - 16	0.000	0.02	100.0	0.044	0.09	92.3	0.040	0.12	92.9	0.049	0.15	91.4
1 - 17	0.000	0.02	100.0	0.045	0.09	92.4	0.043	0.09	92.8	0.050	0.17	91.5
1 - 18	0.000	0.02	100.0	0.046	0.08	92.4	0.044	0.16	92.7	0.051	0.24	91.4
1 - 19	0.000	0.02	100.0	0.047	0.08	92.3	0.045	0.08	92.7	0.052	0.20	91.4
1 - 20	0.000	0.02	100.0	0.048	0.08	92.3	0.048	0.07	92.4	0.054	0.21	91.4
Average		0.02	100.0		0.08	92.4		0.11	92.7		0.19	91.4
1 - 21	0.000	0.02	100.0	0.023	0.06	92.1	0.011	0.10	96.2	0.025	0.19	91.1
1 - 22	0.000	0.02	100.0	0.023	0.06	92.3	0.011	0.12	96.3	0.026	0.20	91.3
1 - 23	0.000	0.02	100.0	0.023	0.06	92.2	0.011	0.10	96.2	0.026	0.17	91.2
1 - 24	0.000	0.03	100.0	0.024	0.07	92.2	0.012	0.09	96.2	0.027	0.25	91.2
1 - 25	0.000	0.02	100.0	0.025	0.06	92.1	0.012	0.09	96.2	0.028	0.21	91.2
Average		0.02	100.0		0.06	92.2		0.10	96.2		0.20	91.2
2 - 26	0.000	0.03	100.0	0.031	0.13	92.1	0.015	0.12	96.3	0.038	0.22	90.2
2 - 27	0.000	0.03	100.0	0.031	0.12	92.2	0.015	0.12	96.2	0.039	0.35	90.4
2 - 28	0.000	0.03	100.0	0.032	0.13	92.1	0.016	0.12	96.1	0.039	0.24	90.3
2 - 29	0.000	0.03	100.0	0.033	0.12	92.0	0.016	0.11	96.1	0.040	0.24	90.2
2 - 30	0.000	0.03	100.0	0.034	0.11	92.1	0.017	0.13	96.0	0.042	0.25	90.2
Average		0.03	100.0		0.12	92.1		0.12	96.1		0.26	90.3
2 - 31	0.000	0.03	100.0	0.031	0.13	92.1	0.015	0.11	96.3	0.038	0.22	90.2
2 - 32	0.000	0.02	100.0	0.031	0.12	92.2	0.015	0.12	96.2	0.039	0.35	90.4
2 - 33	0.000	0.02	100.0	0.032	0.12	92.1	0.016	0.11	96.1	0.039	0.47	90.3
2 - 34	0.000	0.02	100.0	0.033	0.13	92.0	0.016	0.10	96.1	0.040	0.27	90.2
2 - 35	0.000	0.02	100.0	0.034	0.12	92.1	0.017	0.13	96.0	0.042	0.25	90.2
Average		0.02	100.0		0.12	92.1		0.12	96.1		0.31	90.3

We then compare the scenario decomposition Algorithm 1 based on four different MILP formulations (i.e., MILP.B, MILP.VI, MILP.CONV, MILP.VI.CONV) as well as two approximation algorithms (i.e., Algorithm 3, Algorithm 4), and results are reported in Table 3. Compared to directly solving MILP formulations using a solver, as shown in Table 1, we see that for each MILP method, their corresponding scenario decomposition counterpart is more effective and has a much shorter solution time. More precisely, except for the MILP.B model, employing scenario decomposition Algorithm 1 reduces the solution time by about 60% compared to solving its corresponding MILP formulation. This demonstrates the effectiveness of scenario decomposition Algorithm 1. As scenario decomposition based on MILP.VI method took the least time for nearly all the instances, we used it to the accelerated scenario decomposition Algorithm 3.

We see that if the Wasserstein radius θ increases, the solution time does not change too much for all the scenario decomposition methods. As the pandemic factor δ or the number of buses $\boldsymbol{\eta}$ increases, the solution time decreases. As the weight ω increases, the solution time of Algorithm 1 based on MILP.B, MILP.CONV, MILP.VI.CONV decreases, while the solution time of Algorithm 1 based on MILP.VI increases. As the number of routes I increases, the solution time increases for all the scenario decomposition methods.

When implementing the accelerated scenario decomposition Algorithm 3, we forced the lower bound to increase by 1% at each iteration. Therefore, compared to scenario decomposition Algorithm 1 based on MILP.VI, the accelerated scenario decomposition Algorithm 3 significantly reduces the running time by around 80%. Although the accelerated scenario decomposition Algorithm 3 may miss the optimal solution by cutting off plausible solutions, we see from Table 3 that it consistently finds optimal solutions for all the testing cases. We also applied the approximation Algorithm 4 by setting the maximum iteration to be 1000. We notice from Table 3 that for all the testing cases, the approximation Algorithm 4 can obtain an optimal solution and takes within a second to solve. This suggests that the approximation Algorithm 4 can be a strong alternative when the exact methods might not work well. We also see that no matter how the parameters change, the solution time does not change too much for Algorithm 3 and Algorithm 4.

According to the results in Table 1, Table 2, and Table 3, we see that MILP.VI is the most efficient one among four MILP formulations, scenario decomposition Algorithm 1 based on MILP.VI denoted by "Algorithm 1.VI" is the most efficient one among four different scenario decomposition methods using different MILP formulations, and both approximation methods are quite good. To further compare those methods, we applied MILP.B, MILP.VI, SceDecomp.VI, Algorithm 3, and Algorithm 4 to harder instances with a larger objective value. For Algorithm 4, we let the maximum number of iterations equal to 5 and 100 to see the influence of the maximum number of iterations

Table 3 Results of Four Scenario Decomposition Algorithm 1 Based on Four Different MILP Formulations and Two Approximate Methods (i.e., Algorithm 3, Algorithm 4)

Case	Scenario Decomposition								Approximate Methods			
	MILP.B		MILP.VI		MILP.CONV		MILP.VI.CONV		Algorithm 3		Algorithm 4	
	Time	Gap(%)	Time	Gap(%)	Time	Gap(%)	Time	Gap(%)	Time	Gap(%)	Time	Gap(%)
1 - 1	22	0.0	8	0.0	37	0.0	21	0.0	7	0.0	0.06	0.0
1 - 2	14	0.0	6	0.0	28	0.0	16	0.0	8	0.0	0.03	0.0
1 - 3	27	0.0	9	0.0	36	0.0	20	0.0	12	0.0	0.05	0.0
1 - 4	40	0.0	8	0.0	128	0.0	27	0.0	10	0.0	0.04	0.0
1 - 5	72	0.0	8	0.0	83	0.0	71	0.0	11	0.0	0.05	0.0
Average	35	0.0	8	0.0	62	0.0	31	0.0	10	0.0	0.05	0.0
1 - 6	33	0.0	9	0.0	39	0.0	32	0.0	9	0.0	0.04	0.0
1 - 7	19	0.0	7	0.0	35	0.0	21	0.0	7	0.0	0.03	0.0
1 - 8	50	0.0	6	0.0	34	0.0	20	0.0	6	0.0	0.05	0.0
1 - 9	23	0.0	13	0.0	49	0.0	18	0.0	15	0.0	0.04	0.0
1 - 10	22	0.0	17	0.0	69	0.0	38	0.0	10	0.0	0.06	0.0
Average	29	0.0	10	0.0	45	0.0	26	0.0	9	0.0	0.04	0.0
1 - 11	21	0.0	14	0.0	45	0.0	28	0.0	8	0.0	0.04	0.0
1 - 12	17	0.0	11	0.0	38	0.0	18	0.0	7	0.0	0.03	0.0
1 - 13	15	0.0	13	0.0	38	0.0	17	0.0	11	0.0	0.04	0.0
1 - 14	19	0.0	14	0.0	55	0.0	25	0.0	13	0.0	0.04	0.0
1 - 15	49	0.0	13	0.0	60	0.0	49	0.0	11	0.0	0.06	0.0
Average	24	0.0	13	0.0	47	0.0	27	0.0	10	0.0	0.04	0.0
1 - 16	39	0.0	13	0.0	65	0.0	36	0.0	5	0.0	0.03	0.0
1 - 17	25	0.0	11	0.0	52	0.0	33	0.0	5	0.0	0.02	0.0
1 - 18	27	0.0	10	0.0	45	0.0	35	0.0	5	0.0	0.03	0.0
1 - 19	33	0.0	9	0.0	76	0.0	33	0.0	5	0.0	0.02	0.0
1 - 20	33	0.0	9	0.0	54	0.0	31	0.0	3	0.0	0.03	0.0
Average	31	0.0	10	0.0	58	0.0	34	0.0	5	0.0	0.03	0.0
1 - 21	20	0.0	6	0.0	28	0.0	25	0.0	2	0.0	0.03	0.0
1 - 22	14	0.0	6	0.0	26	0.0	18	0.0	4	0.0	0.03	0.0
1 - 23	12	0.0	5	0.0	20	0.0	17	0.0	3	0.0	0.03	0.0
1 - 24	15	0.0	6	0.0	24	0.0	19	0.0	2	0.0	0.03	0.0
1 - 25	13	0.0	7	0.0	26	0.0	17	0.0	2	0.0	0.03	0.0
Average	15	0.0	6	0.0	25	0.0	19	0.0	3	0.0	0.03	0.0
2 - 26	38	0.0	7	0.0	83	0.0	47	0.0	7	0.0	0.04	0.0
2 - 27	44	0.0	7	0.0	51	0.0	29	0.0	9	0.0	0.04	0.0
2 - 28	41	0.0	8	0.0	57	0.0	29	0.0	8	0.0	0.05	0.0
2 - 29	35	0.0	5	0.0	66	0.0	20	0.0	9	0.0	0.03	0.0
2 - 30	25	0.0	12	0.0	53	0.0	24	0.0	9	0.0	0.04	0.0
Average	37	0.0	8	0.0	62	0.0	30	0.0	8	0.0	0.04	0.0
2 - 31	35	0.0	10	0.0	80	0.0	39	0.0	10	0.0	0.04	0.0
2 - 32	34	0.0	7	0.0	49	0.0	24	0.0	6	0.0	0.05	0.0
2 - 33	30	0.0	6	0.0	50	0.0	21	0.0	9	0.0	0.03	0.0
2 - 34	27	0.0	6	0.0	43	0.0	19	0.0	9	0.0	0.05	0.0
2 - 35	21	0.0	9	0.0	53	0.0	22	0.0	6	0.0	0.03	0.0
Average	29	0.0	8	0.0	55	0.0	25	0.0	8	0.0	0.04	0.0

on the solution quality. These two different configurations are denoted by Algorithm 4(max iter 5) and Algorithm 4(max iter 100). Results are displayed in Table 4.

From Table 4, similar to previous results, we see that MILP.VI reduces the solution time compared to MILP.B. Although Algorithm 1.VI outperforms MILP.VI, Algorithm 3 is the best among these three. We also see that Algorithm 4 (max iter 5) has the shortest solution time but misses the optimal solution, while Algorithm 4(max iter 100) takes a longer average solution time. However, its solution quality is better, and its average solution time is overall shorter than that of accelerated scenario decomposition Algorithm 1. Therefore, in practice, we suggest using approximation Algorithm 4 with a properly chosen maximum number of iterations to solve hard instances.

7.2. Results of Larger Instances

This subsection focuses on larger instances. We generated 3 larger instances where the number of bus routes $I \in \{10, 20\}$ and the number of scenarios $N \in \{10, 20\}$. There are $K = 3$ types of buses with nominal capacities $c_1 = 60, c_2 = 80, c_3 = 120$ and each route has 10 to 40 stops. The pandemic factor δ is 0.25. We tested on 20 cases based on these instances using different parameter configu-

Table 4 Results of Instances with a Larger Objective Value

Case	ω	θ	OPT	MILP.B	MILP.VI		Algorithm 1.VI		Algorithm 3		Algorithm 4 (max itr 5)			Algorithm 4(max itr 100)		
				Time	Time	Gap(%)	Time	Gap(%)	Time	Gap(%)	Obj.Val	Time	Gap(%)	Obj.Val	Time	Gap(%)
3 - 36	0.25	0	1.006	56	25	0.0	11	0.0	12	0.0	1.038	1	3.2	1.006	40	0.0
3 - 37		0.25	1.053	33	46	0.0	15	0.0	16	0.0	1.063	3	1.0	1.053	3	0.0
3 - 38		0.5	1.053	19	54	0.0	24	0.0	19	0.0	1.063	3	1.0	1.053	3	0.0
3 - 39		1	1.053	35	31	0.0	39	0.0	23	0.0	1.063	4	1.0	1.053	3	0.0
3 - 40		1.5	1.076	33	12	0.0	23	0.0	41	0.0	1.076	10	0.0	1.076	8	0.0
Average				35	34	0.0	22	0.0	22	0.0		4	1.2		11	0.0
4 - 41	0.5	0	1.032	24	41	0.0	15	0.0	22	0.0	1.056	2	2.3	1.032	56	0.0
4 - 42		0.25	1.105	48	24	0.0	33	0.0	24	0.0	1.135	3	2.7	1.105	5	0.0
4 - 43		0.5	1.105	40	19	0.0	26	0.0	17	0.0	1.135	3	2.7	1.105	5	0.0
4 - 44		1	1.105	23	25	0.0	21	0.0	20	0.0	1.135	4	2.7	1.105	6	0.0
4 - 45		1.5	1.152	35	11	0.0	24	0.0	23	0.0	1.152	10	0.0	1.152	16	0.0
Average				34	24	0.0	24	0.0	21	0.0		4	2.1		18	0.0
5 - 46	1	0	1.085	33	15	0.0	15	0.0	11	0.0	1.093	2	0.8	1.085	65	0.0
5 - 47		0.25	1.210	29	13	0.0	19	0.0	14	0.0	1.231	4	1.7	1.210	3	0.0
5 - 48		0.5	1.210	36	34	0.0	17	0.0	11	0.0	1.231	4	1.7	1.210	3	0.0
5 - 49		1	1.210	25	14	0.0	17	0.0	18	0.0	1.231	4	1.7	1.210	4	0.0
5 - 50		1.5	1.303	36	36	0.0	21	0.0	15	0.0	1.333	19	2.3	1.303	11	0.0
Average				32	22	0.0	18	0.0	14	0.0		7	1.6		17	0.0

rations (i.e., different values of $N, I, \boldsymbol{\eta}, \omega, \theta$). Particularly, cases 51-60, 61-65, 66-70 correspond to instances 6, 7, 8, respectively.

Results of different MILP formulations are reported in Table 5 and Table 6, where "Obj.Val" denotes the best upper bound, "LB" denotes the best lower bound, "Opt.Gap" represents the optimality gap computed as the percentage of the difference between Obj.Val and LB divided by Obj.Val, and "/" denotes the cases that no solution was found within the time limit (i.e., 3600 seconds). Unfortunately, no case can be solved to optimality in time limit. For the cases 64,65,69,70, MILP.VI.CONV cannot even find a feasible solution within the time limit. We notice that MILP.VI consistently obtains the best lower bound among the four MILP formulations almost for each case. However, albeit promising, MILP.VI may not be ideal for solving extremely large instances since the optimality gap is quite large within the time limit.

Table 5 Results of MILP.B and MILP.VI for Solving Larger Instances

Case	N	I	$\boldsymbol{\eta}$	ω	θ	MILP.B					MILP.VI				
						Obj.Val	Time	Opt.Gap(%)	LB	Nodes	Obj.Val	Time	Opt.Gap(%)	LB	Nodes
6 - 51					0	0.743	3600	39.8	0.447	16791	0.743	3600	34.8	0.484	121062
6 - 52					0.25	0.763	3600	48.0	0.397	10684	0.763	3600	37.6	0.476	20532
6 - 53	10	10	(10,15,5)	0.5	0.5	0.775	3600	39.8	0.466	20265	0.775	3600	32.1	0.526	21055
6 - 54					1	0.778	3600	35.1	0.505	14401	0.778	3600	34.3	0.511	143685
6 - 55					1.5	0.817	3600	32.9	0.549	21088	0.817	3600	41.1	0.482	20586
Average							3600	34.0		16646		3600	37.7		65384
6 - 56					0	0.743	3600	24.3	0.563	27659	0.743	3600	34.5	0.487	164536
6 - 57					0.25	0.763	3600	30.5	0.530	25626	0.763	3600	40.8	0.452	20589
6 - 58	10	10	(10,15,5)	1	0.5	0.775	3600	39.4	0.469	18724	0.775	3600	44.7	0.428	20616
6 - 59					1	0.778	3600	35.3	0.503	12208	0.778	3600	35.1	0.505	22077
6 - 60					1.5	0.817	3600	27.5	0.593	34584	0.829	3600	46.9	0.440	20639
Average							3600	31.4		23760		3600	41.0		49691
7 - 61					0	1.309	3600	78.5	0.282	673	1.249	3600	62.4	0.470	13611
7 - 62					0.25	1.311	3600	77.4	0.296	869	1.300	3600	57.8	0.549	1111
7 - 63	10	20	(8,20,8)	0.5	0.5	1.341	3600	78.8	0.285	1040	1.272	3600	56.4	0.555	1116
7 - 64					1	1.316	3600	77.4	0.297	1035	1.340	3600	56.9	0.577	1071
7 - 65					1.5	1.335	3600	76.6	0.313	1058	1.299	3600	51.8	0.626	1085
Average							3600	77.0		935		3600	54.4		3599
8 - 66					0	1.378	3600	81.0	0.261	1	0.798	3600	61.1	0.311	1195
8 - 67					0.25	1.112	3600	80.6	0.215	20691	0.923	3600	64.2	0.331	1052
8 - 68	20	20	(20, 30, 15)	0.5	0.5	1.396	3600	80.5	0.272	1	1.349	3600	75.6	0.329	2958
8 - 69					1	1.332	3600	81.1	0.252	1	1.294	3600	72.9	0.351	1058
8 - 70					1.5	1.353	3600	80.3	0.267	1	0.867	3600	58.7	0.358	1088
Average							3600	80.7		4139		3600	65.8		1470

Table 6 Results of MILP.CONV and MILP.VI.CONV for Solving Larger Instances

Case	MILP.CONV					MILP.VI.CONV				
	Obj.Val	Time	Opt.Gap(%)	LB	Nodes	Obj.Val	Time	Opt.Gap(%)	LB	Nodes
6 - 51	0.755	3600	49.5	0.381	2487	0.755	3600	50.1	0.377	21512
6 - 52	0.773	3600	48.2	0.400	1352	0.773	3600	48.1	0.401	21526
6 - 53	0.787	3600	42.9	0.449	21090	0.775	3600	54.6	0.352	21166
6 - 54	0.788	3600	55.4	0.352	5154	0.778	3600	47.6	0.408	15570
6 - 55	0.861	3600	54.1	0.395	2274	0.829	3600	54.6	0.376	15002
Average		3600	54.8		3714		3600	51.1		15286
6 - 56	0.767	3600	45.4	0.419	20850	0.743	3600	48.6	0.382	21441
6 - 57	0.773	3600	40.5	0.460	1075	0.773	3600	53.5	0.360	15879
6 - 58	0.782	3600	44.9	0.431	1039	0.782	3600	48.5	0.403	20659
6 - 59	0.847	3600	51.3	0.413	1568	0.792	3600	51.2	0.386	18457
6 - 60	0.864	3600	47.8	0.451	21036	0.817	3600	48.8	0.418	12172
Average		3600	49.5		11302		3600	50.0		15315
7 - 61	1.392	3600	65.1	0.485	1	1.443	3600	73.7	0.380	1
7 - 62	1.393	3600	63.6	0.507	1	1.445	3600	69.1	0.446	11
7 - 63	1.385	3600	62.8	0.515	1	1.445	3600	69.0	0.448	7
7 - 64	1.405	3600	65.7	0.482	1	/	3600	/	/	/
7 - 65	1.411	3600	63.1	0.521	2903	/	3600	/	/	/
Average		3600	64.4		1452		3600	70.6		6
8 - 66	1.447	3600	79.1	0.302	1	1.447	3600	96.7	0.048	1
8 - 67	1.446	3600	80.7	0.279	18	1.446	3600	97.3	0.039	1
8 - 68	1.447	3600	75.3	0.357	1	1.447	3600	95.7	0.063	1
8 - 69	1.449	3600	80.3	0.286	1	/	3600	/	/	/
8 - 70	1.412	3600	76.4	0.334	1	/	3600	/	/	/
Average		3600	78.3		1		3600	96.5		1

Results of continuous relaxations of MILP formulations are displayed in Table 7. Similarly, we see that the continuous relaxation of MILP.VI.CONV is the best compared to other formulations; however, it takes the longest time to compute. On the other hand, the continuous relaxation values of MILP.VI are comparable to those of MILP.VI.CONV but takes a much shorter time. This somehow explains why MILP.VI works the best among all the MILP formulations. Results of cases 61-70 show that for very large-scale cases, even solving a continuous relaxation takes a relatively long time to solve to optimality, indicating that exact methods may also have trouble solving these cases to optimality within an hour.

Table 7 Results of Continuous Relaxations of Four Different MILP Formulations for Solving Larger Instances

Case	C.MILP.B		C.MILP.VI		C.MILP.CONV		C.MILP.VI.CONV	
	Obj.Val	Time	Obj.Val	Time	Obj.Val	Time	Obj.Val	Time
6 - 51	0.000	0.16	0.044	0.83	0.030	0.45	0.054	3.53
6 - 52	0.000	0.15	0.045	0.96	0.032	0.44	0.056	2.68
6 - 53	0.000	0.16	0.046	0.58	0.032	0.37	0.056	2.65
6 - 54	0.000	0.34	0.046	1.09	0.033	0.39	0.057	2.72
6 - 55	0.000	0.19	0.048	0.67	0.035	0.40	0.059	3.21
Average		0.20		0.83		0.41		2.96
6 - 56	0.000	0.16	0.044	0.98	0.030	0.36	0.054	2.38
6 - 57	0.000	0.15	0.045	0.74	0.032	0.53	0.056	2.52
6 - 58	0.000	0.14	0.046	0.62	0.032	0.46	0.056	2.53
6 - 59	0.000	0.14	0.046	0.85	0.033	0.62	0.057	3.48
6 - 60	0.000	0.14	0.048	0.79	0.035	0.37	0.059	2.95
Average		0.15		0.80		0.47		2.77
7 - 61	0.000	0.22	0.307	2.45	0.020	1.12	0.307	12.28
7 - 62	0.000	0.24	0.342	1.59	0.021	1.56	0.342	8.82
7 - 63	0.000	0.23	0.352	2.18	0.022	1.58	0.352	8.47
7 - 64	0.000	0.23	0.367	2.35	0.022	1.60	0.367	8.17
7 - 65	0.000	0.22	0.415	2.33	0.023	1.12	0.415	6.76
Average		0.23		2.18		1.40		8.90
8 - 66	0.000	0.52	0.023	5.01	0.003	4.61	0.027	22.80
8 - 67	0.000	0.64	0.024	4.50	0.003	3.69	0.028	25.73
8 - 68	0.000	0.50	0.024	6.05	0.003	3.49	0.028	40.80
8 - 69	0.000	0.56	0.024	5.01	0.003	3.75	0.028	29.88
8 - 70	0.000	0.53	0.025	4.21	0.009	3.12	0.030	29.71
Average		0.55		4.96		3.73		29.79

Results of scenario decomposition Algorithm 1 based on four different MILP formulations, scenario decomposition with grouping Algorithm 2 based on MILP.VI, accelerated scenario decompo-

sition Algorithm 3 based on MILP.VI and approximation algorithm Algorithm 4 (with maximum iteration equal to 5) are shown in Table 8 and Table 9, where "Best.Obj" represents the best objective value obtained by four MILP formulations and Algorithm 1 based on four MILP formulations for each case. In scenario bundling, we used K-means clustering algorithm to group the scenarios into 5 groups for instance 6 and 7, 10 groups for instance 8. It is seen that Algorithm 4 is the only method that can find feasible solutions to cases 61-65. For cases 66-70, only Algorithm 1 based on the MILP.VI formulation, Algorithm 3, and Algorithm 4 can obtain feasible solutions within the time limit. For the hard cases 61-70, Algorithm 4 finds better solutions than the best ones output by the exact methods. We notice that, in Table 9, based on the MILP.VI formulation, compared with Algorithm 1, Algorithm 2 using scenario group improves the optimality gap for instance 6, which shows that grouping can help solve medium-size instances. However, Algorithm 2 cannot obtain any feasible solution for instance 8 within the time limit, while Algorithm 1 can find at least one solution. This may be because scenario-grouping subproblems are more challenging to solve than those without grouping. Thus, for the large-scale instances, we suggest using the naive scenario decomposition method (i.e., Algorithm 1). For easy cases 51-60, both Algorithm 3 and Algorithm 4 find exactly the same solutions as the best ones output by the exact methods. Overall, we conclude that among all the exact methods, when solving moderate-sized cases, Algorithm 1 based on MILP.VI outperforms other methods by providing the smallest optimality gaps; on the other hand, for the hard cases, none of the exact methods works well, MILP.VI and Algorithm 1 based on MILP.VI are slightly better than others since they can consistently find a better solution. We also see that Algorithm 4 consistently finds either a better solution or the same quality solution. Therefore, in practice, when facing very large-scale cases or involving multiple-round cross-validations, we suggest using Algorithm 4.

7.3. Out-of-sample Performance

In this section, we tested the out-of-sample performance of the proposed DrFRAM model. All the instances in this subsection are relatively small and were solved to optimality by MILP.VI. Let $\mathbb{P}^\infty$ denote the sample probability distribution, and $\mathbb{P}^*$ be the true distribution of random parameters $\tilde{\boldsymbol{\xi}}$. Motivated by Esfahani and Kuhn (2018), we define the out-of-sample probability as

$$\mathbb{P}^\infty \left\{ \tilde{\boldsymbol{\xi}} : v^* < \mathbb{E}_{\mathbb{P}^*} \left[\mathcal{Q}\left(\boldsymbol{n}^D, \boldsymbol{x}^D, \tilde{\boldsymbol{\xi}} \right) \right] + \rho \mathrm{CVaR}_{1-\varepsilon} \left(\boldsymbol{n}^D, \boldsymbol{x}^D, \tilde{\boldsymbol{\xi}} \right) \right\},$$

where $(\boldsymbol{n}^D, \boldsymbol{x}^D)$ denotes the optimal solution of the DrFRAM model and $\mathcal{Q}(\cdot,\cdot,\cdot)$ represents the recourse function. That is, we would like to ensure that the probability that the DrFRAM optimal value is smaller than the mean-risk objective is small, e.g., no larger than 2α; in our numerical study, we let $\alpha = 5\%$, $\rho = 0.1$ and $\varepsilon = 0.1$. To measure this out-of-sample probability, for any given

Table 8 Results of Three Scenario Decomposition Algorithm 1 Based on Three Different MILP Formulations (i.e., MILP.B, MILP.CONV, MILP.CONV.VI) for Solving Larger Instances

Case	Scenario Decomposition											
	MILP.B				MILP.CONV				MILP.VI.CONV			
	Obj.Val	LB	Time	Opt.Gap(%)	Obj.Val	LB	Time	Opt.Gap(%)	Obj.Val	LB	Time	Opt.Gap(%)
6 - 51	0.743	0.708	3600	4.7	0.743	0.701	3600	5.7	0.743	0.717	3600	3.6
6 - 52	0.763	0.734	3600	3.8	0.763	0.731	3600	4.3	0.763	0.743	3600	2.7
6 - 53	0.775	0.745	3600	3.9	0.775	0.738	3600	4.7	0.775	0.740	3600	4.4
6 - 54	0.778	0.748	3600	3.8	0.778	0.741	3600	4.7	0.778	0.756	3600	2.8
6 - 55	0.817	0.785	3600	3.9	0.817	0.783	3600	4.2	0.817	0.795	3600	2.8
Average			3600	4.0			3600	4.7			3600	3.3
6 - 56	0.743	0.717	3600	3.6	0.743	0.702	3600	5.6	0.743	0.717	3600	3.6
6 - 57	0.763	0.737	3600	3.4	0.763	0.732	3600	4.1	0.763	0.743	3600	2.7
6 - 58	0.775	0.745	3600	3.9	0.775	0.738	3600	4.7	0.775	0.740	3600	4.4
6 - 59	0.778	0.748	3600	3.8	0.778	0.741	3600	4.7	0.778	0.748	3600	3.8
6 - 60	0.817	0.785	3600	3.9	0.817	0.783	3600	4.2	0.817	0.795	3600	2.8
Average			3600	3.7			3600	4.7			3600	3.5
7 - 61	/	/	3600	/	/	/	3600	/	/	/	3600	/
7 - 62	/	/	3600	/	/	/	3600	/	/	/	3600	/
7 - 63	/	/	3600	/	/	/	3600	/	/	/	3600	/
7 - 64	/	/	3600	/	/	/	3600	/	/	/	3600	/
7 - 65	/	/	3600	/	/	/	3600	/	/	/	3600	/
Average			3600	/			3600	/			3600	/
8 - 66	/	/	3600	/	/	/	3600	/	/	/	3600	/
8 - 67	/	/	3600	/	/	/	3600	/	/	/	3600	/
8 - 68	/	/	3600	/	/	/	3600	/	/	/	3600	/
8 - 69	/	/	3600	/	/	/	3600	/	/	/	3600	/
8 - 70	/	/	3600	/	/	/	3600	/	/	/	3600	/
Average			3600	/			3600	/			3600	/

Table 9 Results of the Scenario Decomposition Algorithm 1 Based on MILP.VI, Denoted by Algorithm 1.VI, Scenario Decomposition with Grouping Algorithm 2 Based on MILP.VI, Denoted by Algorithm 2.VI, and Two Approximate Methods (i.e., Algorithm 3, Algorithm 4) for Solving Larger Instances

Case	Algorithm 1.VI				Algorithm 2.VI				Best.Obj	Algorithm 3		Algorithm 4	
	Obj.Val	LB	Time	Opt.Gap(%)	Obj.Val	LB	Time	Opt.Gap(%)		Obj.Val	Time	Obj.Val	Time
6 - 51	0.743	0.719	3600	3.3	0.743	0.719	3600	3.3	0.743	0.743	119	0.743	0.01
6 - 52	0.763	0.748	3600	2.0	0.763	0.750	3600	1.8	0.763	0.763	102	0.763	0.02
6 - 53	0.775	0.757	3600	2.2	0.775	0.759	3600	2.1	0.775	0.775	152	0.775	1.00
6 - 54	0.778	0.762	3600	2.0	0.778	0.765	3600	1.7	0.778	0.778	123	0.778	0.01
6 - 55	0.817	0.797	3600	2.5	0.817	0.798	3600	2.5	0.817	0.817	171	0.817	0.02
Average			3600	2.4			3600	2.3			133		0.21
6 - 56	0.743	0.719	3600	3.3	0.743	0.719	3600	3.3	0.743	0.743	159	0.743	0.01
6 - 57	0.763	0.748	3600	2.0	0.763	0.750	3600	1.8	0.763	0.763	139	0.763	0.02
6 - 58	0.775	0.754	3600	2.7	0.775	0.759	3600	2.1	0.775	0.775	233	0.775	1.00
6 - 59	0.778	0.766	3600	1.5	0.778	0.765	3600	1.7	0.778	0.778	234	0.778	0.01
6 - 60	0.817	0.800	3600	2.1	0.817	0.800	3600	2.1	0.817	0.817	212	0.817	0.02
Average			3600	2.3			3600	2.2			195		0.21
7 - 61	/	/	3600	/	/	/	3600	/	1.249	/	3600	1.124	3600
7 - 62	/	/	3600	/	/	/	3600	/	1.300	/	3600	1.139	3600
7 - 63	/	/	3600	/	/	/	3600	/	1.272	/	3600	1.140	3600
7 - 64	/	/	3600	/	/	/	3600	/	1.316	/	3600	1.148	3600
7 - 65	/	/	3600	/	/	/	3600	/	1.299	/	3600	1.163	3600
Average			3600	/			3600	/			3600		3600
8 - 66	0.768	0.695	3600	9.4	/	/	3600	/	0.798	0.767	3600	0.764	776
8 - 67	0.789	0.718	3600	9.0	/	/	3600	/	0.923	0.789	3600	0.780	270
8 - 68	0.800	0.723	3600	9.7	/	/	3600	/	1.349	0.800	3600	0.789	390
8 - 69	0.819	0.735	3600	10.3	/	/	3600	/	1.294	0.808	3600	0.800	349
8 - 70	0.849	0.769	3600	9.4	/	/	3600	/	0.867	0.848	3600	0.839	986
Average			3600	9.6			3600	/			3600		554

θ, we first solved the DrFRAM model using generated data and then generated new samples with the same sample size and obtained $(1-\alpha)$ confidence intervals of the objective value by plugging in the solution to the DrFRAM and the mean-risk models. We repeated the same procedure for a list of θ values and selected the smallest θ that the confidence interval for the DrFRAM objective value is beyond that of the mean-risk objective value, i.e., to guarantee out-of-sample probability to be at most 2α.

In our numerical study, we supposed that the number of bus routes I is 5, each route has 10 to 40 stops, there are $K=3$ types of buses with nominal capacities $c_1=60, c_2=80, c_3=120$, the

number of buses is $\boldsymbol{\eta} = (6, 9, 3)$, the pandemic factor δ is 0.25, the weight ω is 0.5, and the number of scenarios $N = 10$. We also assumed that the passenger arrival rates follow discrete uniform distributions between 0 and 25. The percentage of passengers alighting from the bus follows a triangular distribution with a lower limit of 0.2, an upper limit of 0.9, and a mode of 0.4. To select the smallest θ that guaranteed a 90% in-sample performance, we adopted the following procedure: (i) for each $\theta \in \{0, 0.2, \ldots, 3\}$, generate 10 scenarios for the DrFRAM model; (ii) generate 10 scenarios by plugging the solution from part (i) into the DrFRAM and mean-risk models; (iii) repeat part (i) and (ii) for 15 times and derive the asymptotic 95% confidence intervals of both models; and (iv) Choose the smallest θ such that the confidence interval of the DrFRAM model is completely above the confidence interval of the mean-risk model. We also compared the performances of the SAA model (i.e., in DrFRAM (5) with $\theta = 0$) and the DrFRAM model with the following procedure: (i) generate 10 scenarios to solve the DrFRAM model with the best tuned Wasserstein radius and the SAA model separately; (ii) generate 10 scenarios with the same parameters and evaluate the DrFRAM and SAA solutions using the SAA model; (iii) repeat part (ii) 1000 times and compute the asymptotic 95% confidence interval for the mean of the objective value.

The result is displayed in Figure 1, and the best Wasserstein radius is $\theta = 1.6$. By comparing the results of the SAA model and the DrFRAM model with $\theta = 1.6$, we see that the SAA solution yields a higher objective value than the DrFRAM solution. This demonstrates that the DrFRAM model with the best tuned Wasserstein radius can outperform the SAA model when there are very limited data.

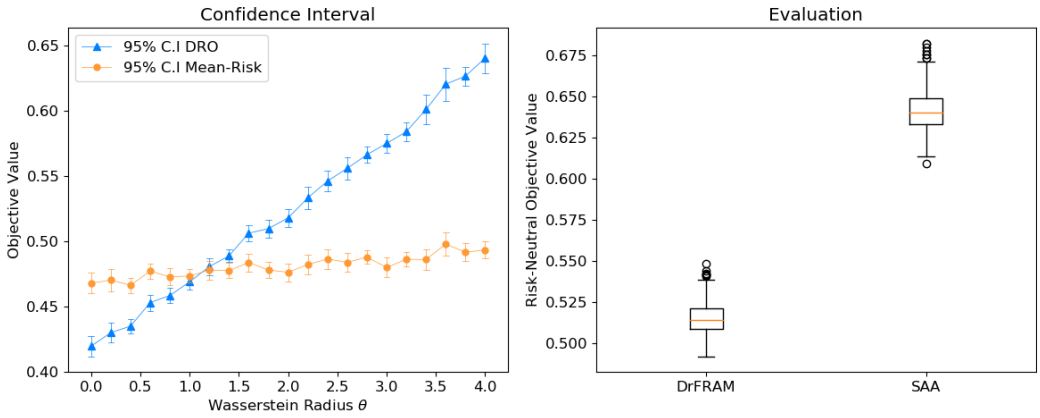

Figure 1 **Tuning the Wasserstein Radius and Comparing the DrFRAM and SAA Solutions.**

7.4. Blacksburg Transit Case Study

In this section, we applied approximation Algorithm 4 to solve the real-world Blacksburg Transit allocation problem since the dataset provided by Blacksburg Transit is of a similar scale as the largest cases studied in the previous section. The operation data was provided by Blacksburg Transit, and passenger arrival and alighting data were collected in September 2020, amid the COVID-19 pandemic. Blacksburg Transit has 17 routes, three types of buses with different capacities, where the number of buses with the nominal capacity equal to 60, 80, 120 are $8, 20, 8$, respectively, and each route has 10 to 40 stops. As public policies and government regulations change as the pandemic evolves, we study the pandemic factor δ varying from 0.25 (social distancing required by CDC) to 1 (fully operated). We set the weight $\omega = 0.5$ in this numerical study since the priority during a pandemic is to enforce social distancing. We adopted the following cross-validation procedure to choose the best Wasserstein radius θ for a given δ using 22 scenarios as training data: (i) randomly select 16 scenarios for solving the model; (ii) for each $\theta \in \{0, 0.2, \ldots, 2\}$, solve the DrFRAM model using the 16 selected scenarios and evaluate the solution using the remaining 6 scenarios; (iii) repeat (i) and (ii) for 15 times and derive the asymptotic 95% confidence interval of the objective values. We followed the same procedure in the previous subsection to select the smallest θ such that the confidence interval for the DrFRAM objective value is beyond that of the mean-risk objective value. We let $\rho = 0.5$ and $\varepsilon = 0.1$. A time limit of 3600 seconds was set for a single run. Note that when all the DrFRAM objective values corresponding to different Wasserstein radii θ are much less than 1, then the cross-validation takes about 1000 seconds; otherwise, if for some Wasserstein radii, their corresponding DrFRAM objective values are around or exceeding 1, then the cross-validation can take up to 15 hours.

We see that for any $\delta \in \{0.5, 0.75, 1\}$ and for any $\theta \in \{0, 0.2, \ldots, 6\}$, since objective values for all the scenarios are observed to be less than 1, according to Proposition 8 and its remark, the solution obtained by approximation Algorithm 4 can be very close to the optimality. Besides, since we obtain the same solution when $\delta \geq 0.5$, we only consider $\delta \in \{0.25, 0.5\}$. Thus, according to our cross-validation results in Table 10 and Figure 2, we chose $\theta = 1.8$ as the best Wasserstein radius when $\delta = 0.25$ and $\theta = 2.6$ for $\delta = 0.5$.

We evaluated the current bus assignment from Blacksburg Transit and compared the results of the SAA model and DrFRAM model with the best tuned Wasserstein radius θ with unused 6 scenarios as testing data. To simulate the possible changes in data, we applied a truncated Gaussian noise to each testing sample. The noise for the proportion of passengers alighting from the bus follows $\mathcal{N}(0, 0.001t)$ truncated to be nonnegative, while the noise for the passenger arrival rate follows $\mathcal{N}(0, 0.1t)$ rounded to the nearest nonnegative integer, where the parameter $t \in \{0, \ldots, 50\}$. We repeated the sampling process 100 times to derive the 95% asymptotic confidence interval.

Table 10 Cross Validation Results of Blacksburg Transit Data

$\delta = 0.25$			$\delta = 0.5$		
θ	95% C.I. of DrFRAM	95% C.I. of Mean-Risk	θ	95% C.I. of DrFRAM	95% C.I. of Mean-Risk
0.0	[0.122 , 0.135]	[0.219 , 0.254]	0.0	[0.056 , 0.059]	[0.104 , 0.110]
0.2	[0.559 , 0.580]	[0.555 , 0.590]	0.2	[0.290 , 0.296]	[0.283 , 0.300]
0.4	[0.565 , 0.586]	[0.559 , 0.594]	0.4	[0.292 , 0.298]	[0.283 , 0.300]
0.6	[0.571 , 0.593]	[0.563 , 0.598]	0.6	[0.295 , 0.302]	[0.285 , 0.302]
0.8	[0.578 , 0.600]	[0.569 , 0.603]	0.8	[0.299 , 0.306]	[0.287 , 0.304]
1.0	[0.588 , 0.612]	[0.578 , 0.616]	1.0	[0.304 , 0.311]	[0.291 , 0.310]
1.2	[1.051 , 1.131]	[0.845 , 1.001]	1.2	[0.651 , 0.741]	[0.535 , 0.712]
1.4	[1.130 , 1.217]	[1.005 , 1.129]	1.4	[0.797 , 0.952]	[0.747 , 0.996]
1.6	[1.134 , 1.224]	[1.001 , 1.163]	1.6	[0.686 , 0.774]	[0.544 , 0.656]
1.8	[1.073 , 1.154]	[0.865 , 0.984]	1.8	[0.690 , 0.781]	[0.550 , 0.670]
2.0	[1.016 , 1.075]	[0.784 , 0.873]	2.0	[0.793 , 0.929]	[0.663 , 0.820]
2.2	[1.412 , 1.452]	[1.081 , 1.090]	2.2	[0.996 , 1.090]	[0.820 , 1.023]
2.4	[1.418 , 1.451]	[1.070 , 1.082]	2.4	[1.032 , 1.136]	[0.847 , 1.027]
2.6	[1.377 , 1.418]	[1.099 , 1.149]	2.6	[1.038 , 1.130]	[0.824 , 0.956]
2.8	[1.427 , 1.454]	[1.083 , 1.089]	2.8	[1.044 , 1.145]	[0.831 , 0.989]
3.0	[1.430 , 1.462]	[1.084 , 1.098]	3.0	[1.151 , 1.264]	[0.968 , 1.138]

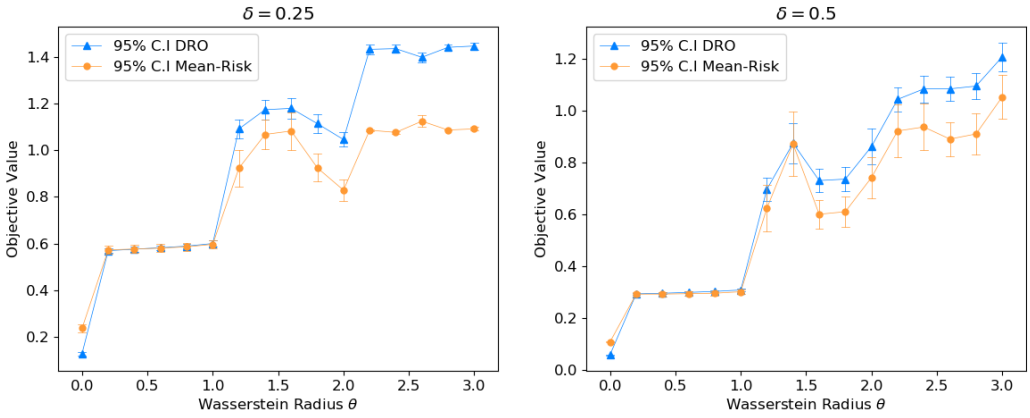

Figure 2 Confidence Intervals of DrFRAM Objective Values and Mean-Risk Objective Values

The result is illustrated in Figure 3. It is seen that our solutions can significantly reduce the highest bus utilization rate and passenger abandon rate compared to the current assignment from Blacksburg Transit. We also see that our DrFRAM solution outperforms SAA one across all the noise levels. This indicates that our proposed approach can consistently provide a better solution than SAA even when the underlying distribution of uncertain parameters is unknown or varies over time.

We analyzed the impact of model parameters, i.e., pandemic factor δ and weight ω by varying δ from 0.25 to 1 and ω from 0.25 to 2. While varying δ, we let $\omega = 0.5$; while varying ω, we let $\delta = 0.25$. We tuned θ following the same procedure as introduced at the beginning of this subsection for the different parameter combinations and did the evaluation with the best-tuned θ to derive 95% confidence intervals. The results are illustrated in Figure 4. It is seen that the objective value

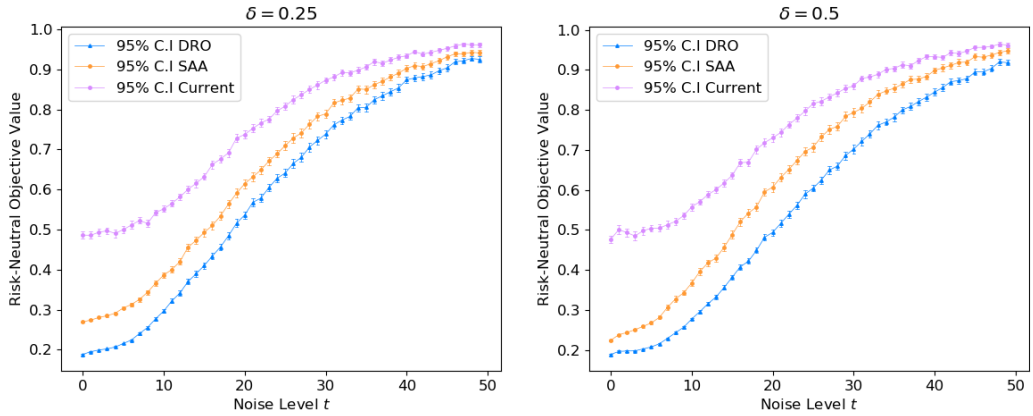

Figure 3 **A Comparison of Risk-Neutral Objective Values for Three Different Methods with Different Noise Levels**

gets smaller when the pandemic factor δ increases, which indicates that the weighted sum of bus utilization and passenger abandon rates decreases when more passengers are allowed to get on board. The objective value increases when the weight ω increases since a higher weight ω leads to a larger objective value.

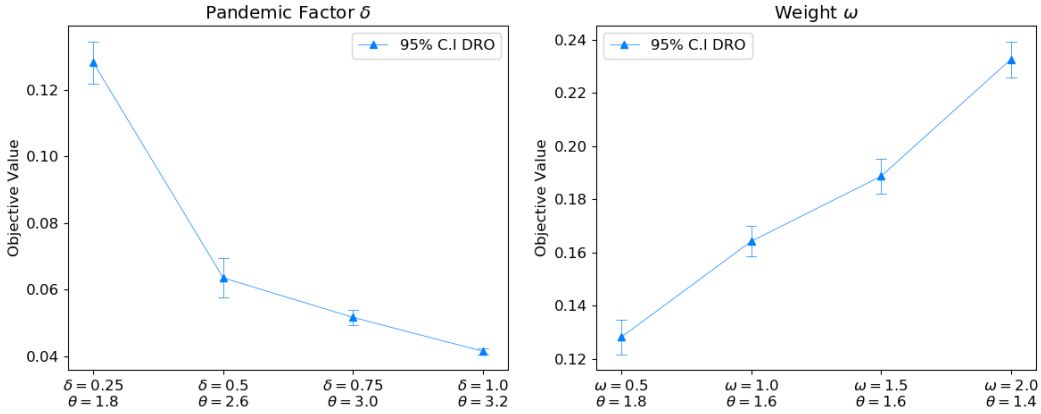

Figure 4 **Sensitivity Analysis on Model Parameters**

We also compared the weighted sum of bus utilization rate and passenger abandon rate among all the routes, as illustrated in Figure 5. It is seen that our solution yields much stable route-based objectives than the current one from Blacksburg Transit or SAA one. This implies that our DrFRAM solution can be indeed fairer and significantly reduce the transit resource inequity among different routes. We also compared our proposed assignment solution with the current one in average total bus capacity for each route, as illustrated in Figure 6. It is seen that our proposed solution assigns more buses to several routes but fewer buses to the others than the current assignment.

For example, our solution assigns more buses to route MSA (see Figure 7), connecting downtown Blacksburg to the campus. Although passenger demand decreased due to the COVID-19 pandemic, residents still commuted to downtown for daily essential products. On the other hand, our solution assigns fewer buses to route HWA and HWB, which connect the largest residential communities to the campus, since most residents were taking virtual classes and working from home. By exploring the outdoor activities, we see that passengers reduced daily commute to school or work but kept the necessary short trips for daily essentials during the pandemic, which could result in more imbalance in passenger demand among different routes. Our data-driven model can help enhance the policy-making for Blacksburg Transit by handling the shifts in passenger demand among bus routes and stops during a pandemic.

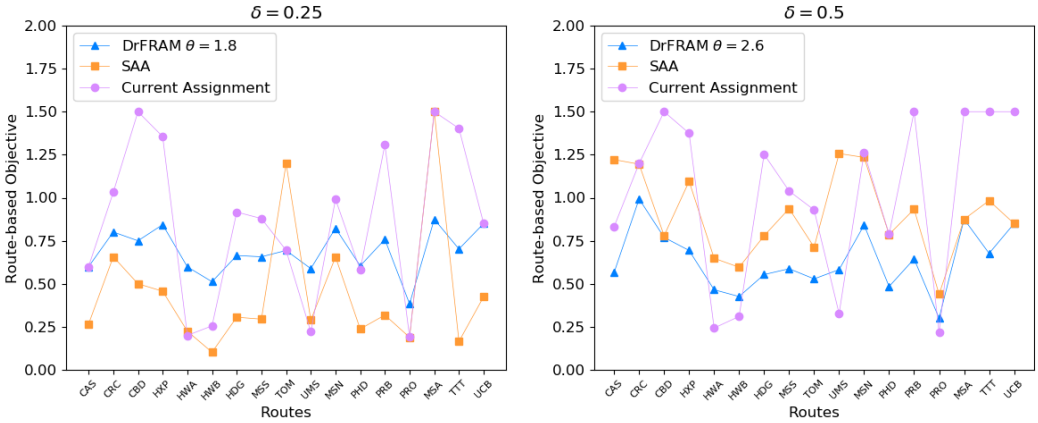

Figure 5 **A Comparison of Route-based Weighted Sum of Bus Utilization Rate and Passenger Abandon Rate Among Three Different Results. Here, xticks are route names of Blacksburg Transit and route-based objective represents the weighted sum of bus utilization rate and passenger abandon rate for each route.**

8. Conclusion

In this paper, we study the transit resource allocation problem to minimize the highest utilization rate and the largest passenger abandon rate under stochastic passenger arrival and alighting rates. We propose a DrFRAM under type-∞ Wasserstein ambiguity set, which is proven to be NP-hard. To simplify the DrFRAM, we derive the monotonicity properties of the DrFRAM and use McCormick inequalities to linearize the nonlinear components, which allows us to derive an MILP formulation. To further improve the MILP formulation, valid inequalities and stronger formulations are derived. We also develop scenario decomposition methods and No-one-left based approximation algorithm to solve DrFRAM. Finally, we numerically demonstrate the effectiveness of the proposed approaches in both small and large instances and apply them to solve the real-world instance using

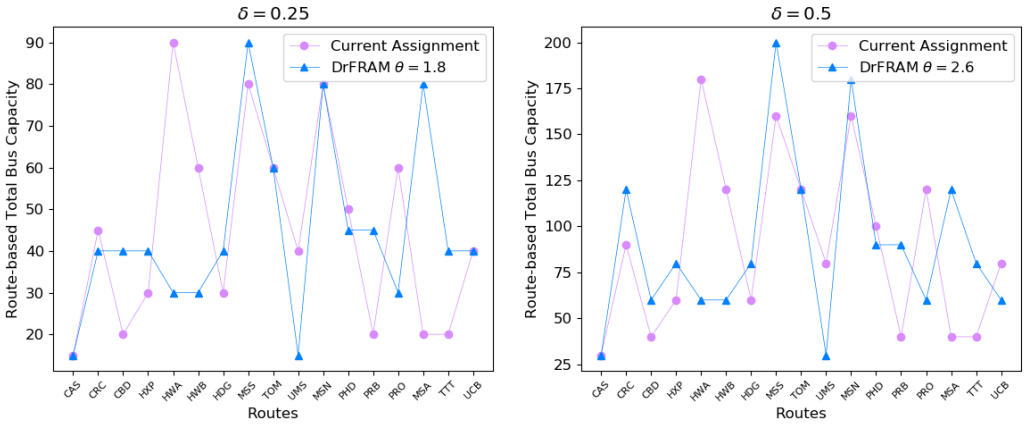

Figure 6 **A Comparison of Routed-based Total Bus Capacity Between the Current Assignment and DrFRAM Result**

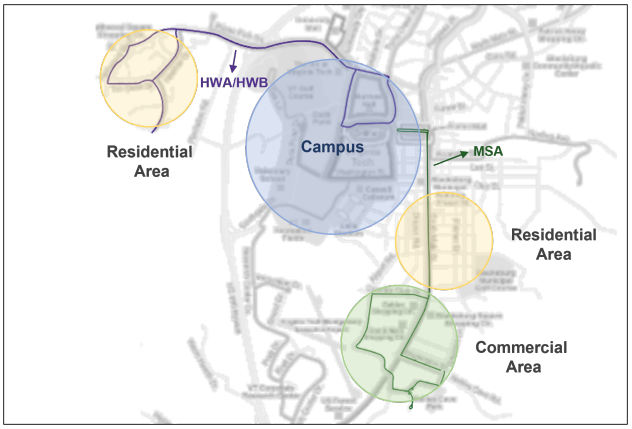

Figure 7 **Illustration of Route HWA/HWB and Route MSA**

the data provided by Blacksburg Transit. Compared to the current allocation plan, our result is demonstrated to be more robust and fairer. More importantly, Blacksburg Transit is using our model to help measure the unbalance of their transit resource assignment. Blacksburg Transit is now following our recommendations in the internal system to make better decisions during the COVID-19 crisis. The proposed DrFRAM framework can be generalized to other problems with uncertain demand and fair resource allocation issues. For instance, we can extend our framework to the truck platooning problems with fair truck resource allocation and passenger demand uncertainty. We can also generalize the model to the call center problems, aiming to minimize the waiting time and maximize fairness and customer satisfaction.

9. Acknowledgments

This research has been supported by the National Science Foundation grant 2046426 and 2153607. The authors benefited from the efforts of undergraduate researchers Kunpeng Liu, Mark Seymour,

Yucheng Gao, and Shangzheng Ji in the early stages of this research. The authors also gratefully acknowledge the helpful comments from the editors and two anonymous reviewers.

References

Carlos F Daganzo and Yanfeng Ouyang. *Public transportation systems: Principles of system design, operations planning and real-time control*. World Scientific, 2010.

Avishai Ceder. *Public transit planning and operation: Modeling, practice and behavior*. CRC press, 2016.

Steven Chien and Paul Schonfeld. Optimization of grid transit system in heterogeneous urban environment. *Journal of Transportation Engineering*, 123(1):28–35, 1997.

Joaquin De Cea and Enrique Fernández. Transit assignment for congested public transport systems: an equilibrium model. *Transportation science*, 27(2):133–147, 1993.

Luca Quadrifoglio, Randolph W Hall, and Maged M Dessouky. Performance and design of mobility allowance shuttle transit services: bounds on the maximum longitudinal velocity. *Transportation science*, 40(3): 351–363, 2006.

JF Guan, Hai Yang, and Sumedha Chandana Wirasinghe. Simultaneous optimization of transit line configuration and passenger line assignment. *Transportation Research Part B: Methodological*, 40(10):885–902, 2006.

Mansoureh Jeihani, Lei Zhang, Anam Ardeshiri, Arash Amiri, Arefeh Nasri, Kiana Roshan Zamir, Babak Baghaei, et al. Development of a framework for transit-oriented development (tod). Technical report, Maryland. State Highway Administration, 2013.

Valérie Guihaire and Jin-Kao Hao. Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, 42(10):1251–1273, 2008.

Daniel Hörcher and Alejandro Tirachini. A review of public transport economics. *Economics of Transportation*, 25:100196, 2021.

Seyed Mohammad Nourbakhsh and Yanfeng Ouyang. A structured flexible transit system for low demand areas. *Transportation Research Part B: Methodological*, 46(1):204–216, 2012.

Yanfeng Ouyang, Seyed Mohammad Nourbakhsh, and Michael J Cassidy. Continuum approximation approach to bus network design under spatially heterogeneous demand. *Transportation Research Part B: Methodological*, 68:333–344, 2014.

Peng Will Chen and Yu Marco Nie. Connecting e-hailing to mass transit platform: Analysis of relative spatial position. *Transportation Research Part C: Emerging Technologies*, 77:444–461, 2017.

Christina Iliopoulou and Konstantinos Kepaptsoglou. Integrated transit route network design and infrastructure planning for on-line electric vehicles. *Transportation Research Part D: Transport and Environment*, 77:178–197, 2019. ISSN 1361-9209. doi: https://doi.org/10.1016/j.trd.2019.10.016. URL `https://www.sciencedirect.com/science/article/pii/S1361920919305000`.

Mojtaba Abdolmaleki, Neda Masoud, and Yafeng Yin. Transit timetable synchronization for transfer time minimization. *Transportation Research Part B: Methodological*, 131:143–159, 2020.

Azucena Román-De la Sancha, Juan M Mayoral, and Luis I Román. A novel approach for transit transfer stations design optimization in densely populated cities. *Procedia computer science*, 130:1013–1018, 2018.

Timothy C Matisziw, Alan T Murray, and Changjoo Kim. Strategic route extension in transit networks. *European journal of operational research*, 171(2):661–673, 2006.

Jian Gang Jin, Kwong Meng Teo, and Amedeo R Odoni. Optimizing bus bridging services in response to disruptions of urban transit rail networks. *Transportation Science*, 50(3):790–804, 2016.

Richa Agarwal and Özlem Ergun. Ship scheduling and network design for cargo routing in liner shipping. *Transportation Science*, 42(2):175–196, 2008.

Arthur Mahéo, Philip Kilby, and Pascal Van Hentenryck. Benders decomposition for the design of a hub and shuttle public transit system. *Transportation Science*, 53(1):77–88, 2019.

Zhi-Chun Li, William HK Lam, and SC Wong. The optimal transit fare structure under different market regimes with uncertainty in the network. *Networks and Spatial Economics*, 9(2):191–216, 2009.

Michael Fernandes, Logan Walls, Sean Munson, Jessica Hullman, and Matthew Kay. Uncertainty displays using quantile dotplots or cdfs improve transit decision-making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

Zhi-Chun Li, William HK Lam, and Agachai Sumalee. Modeling impact of transit operator fleet size under various market regimes with uncertainty in network. *Transportation Research Record*, 2063(1):18–27, 2008.

Xiao Fu and William HK Lam. A network equilibrium approach for modelling activity-travel pattern scheduling problems in multi-modal transit networks with uncertainty. *Transportation*, 41(1):37–55, 2014.

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

Yuval Hadas and Matan Shnaiderman. Public-transit frequency setting using minimum-cost approach with stochastic demand and travel time. *Transportation Research Part B: Methodological*, 46(8):1068–1084, 2012.

Ashish Kulshrestha, Yingyan Lou, and Yafeng Yin. Pick-up locations and bus allocation for transit-based evacuation planning with demand uncertainty. *Journal of Advanced Transportation*, 48(7):721–733, 2014.

Kun An and Hong K Lo. Two-phase stochastic program for transit network design under demand uncertainty. *Transportation Research Part B: Methodological*, 84:157–181, 2016.

Jinpeng Liang, Jianjun Wu, Ziyou Gao, Huijun Sun, Xin Yang, and Hong K Lo. Bus transit network design with uncertainties on the basis of a metro network: A two-step model framework. *Transportation Research Part B: Methodological*, 126:115–138, 2019.

Gyugeun Yoon and Joseph YJ Chow. Contextual bandit-based sequential transit route design under demand uncertainty. *Transportation Research Record*, 2674(5):613–625, 2020.

Karmel S Shehadeh. A distributionally robust optimization approach for a stochastic mobile facility routing and scheduling problem. *arXiv preprint arXiv:2009.10894*, 2020.

Luyu Liu, Harvey J Miller, and Jonathan Scheff. The impacts of covid-19 pandemic on public transit demand in the united states. *Plos one*, 15(11):e0242476, 2020.

Baichuan Mo, Kairui Feng, Yu Shen, Clarence Tam, Daqing Li, Yafeng Yin, and Jinhua Zhao. Modeling epidemic spreading through public transit using time-varying encounter network. *Transportation Research Part C: Emerging Technologies*, 122:102893, 2021.

Gongyu Chen, Xinyu Fei, Huiwen Jia, Xian Yu, and Siqian Shen. An optimization-and-simulation framework for redesigning university campus bus system with social distancing. *arXiv preprint arXiv:2010.10630*, 2020.

Konstantinos Gkiotsalitis and Oded Cats. Optimal frequency setting of metro services in the age of covid-19 distancing measures. *Transportmetrica A: Transport Science*, pages 1–23, 2021.

Hongyuan Yang and Marco Nie. Optimizing operational strategies for mass transit systems in response to a global pandemic. *Available at SSRN 3721007*, 2020.

Daniel Kuhn, Peyman Mohajerin Esfahani, Viet Anh Nguyen, and Soroosh Shafieezadeh-Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine learning. In *Operations Research & Management Science in the Age of Analytics*, pages 130–166. INFORMS, 2019.

Bozidar Radunovic and Jean-Yves Le Boudec. A unified framework for max-min and min-max fairness with applications. *IEEE/ACM Transactions on networking*, 15(5):1073–1083, 2007.

Jianbo Du, Liqiang Zhao, Jie Feng, and Xiaoli Chu. Computation offloading and resource allocation in mixed fog/cloud computing systems with min-max fairness guarantee. *IEEE Transactions on Communications*, 66(4):1594–1608, 2017.

Rui Gao and Anton J Kleywegt. Distributionally robust stochastic optimization with Wasserstein distance. *arXiv preprint arXiv:1604.02199*, 2016.

Jose Blanchet, Karthyek Murthy, and Nian Si. Confidence regions in Wasserstein distributionally robust estimation. *arXiv preprint arXiv:1906.01614*, 2019a.

Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166, 2018.

Jose Blanchet and Karthyek Murthy. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600, 2019.

Grani A Hanasusanto and Daniel Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Operations Research*, 66(3):849–869, 2018.

Zhi Chen, Daniel Kuhn, and Wolfram Wiesemann. Data-driven chance constrained programs over Wasserstein balls. *arXiv preprint arXiv:1809.00210*, 2018.

Weijun Xie. On distributionally robust chance constrained programs with Wasserstein distance. *Mathematical Programming*, pages 1–41, 2019.

Soroosh Shafieezadeh Abadeh, Viet Anh Nguyen, Daniel Kuhn, and Peyman Mohajerin Mohajerin Esfahani. Wasserstein distributionally robust kalman filtering. In *Advances in Neural Information Processing Systems*, pages 8474–8483, 2018.

Jose Blanchet, Yang Kang, and Karthyek Murthy. Robust Wasserstein profile inference and applications to machine learning. *Journal of Applied Probability*, 56(3):830–857, 2019b.

Zhi Chen and Weijun Xie. Sharing the value-at-risk under distributional ambiguity. *Available at SSRN 3400033*, 2019.

Dimitris Bertsimas, Shimrit Shtern, and Bradley Sturt. A data-driven approach for multi-stage linear optimization. *Available at Optimization Online*, 2018.

Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.

Nicolás Garcia Trillos and Dejan Slepčev. On the rate of convergence of empirical measures in ∞ transportation distance. *Canadian Journal of Mathematics*, 67(6):1358–1383, 2015.

Weijun Xie. Tractable reformulations of two-stage distributionally robust linear programs over the type-∞ Wasserstein ball. *Operations Research Letters*, 48(4):513–523, 2020.

Weijun Xie, Jie Zhang, and Shabbir Ahmed. Distributionally robust bottleneck combinatorial problems: Uncertainty quantification and robust decision making. *Mathematical Programming*, pages 1–44, 2021.

Garth P McCormick. Computability of global solutions to factorable nonconvex programs: Part I- convex underestimating problems. *Mathematical programming*, 10(1):147–175, 1976.

Egon Balas. Disjunctive programming. *Annals of discrete mathematics*, 5:3–51, 1979.

Shabbir Ahmed. A scenario decomposition algorithm for 0–1 stochastic programs. *Operations Research Letters*, 41(6):565–569, 2013.

Akshay Gupte, Shabbir Ahmed, Myun Seok Cheon, and Santanu Dey. Solving mixed integer bilinear problems using milp formulations. *SIAM Journal on Optimization*, 23(2):721–744, 2013.

Appendix A: Proofs

A.1. Proof of Proposition 1

Proposition 1 *Solving DrFRAM (5) is NP-hard even when* $N=1, K=2, \theta=0$.

Proof: Let us consider the following NP-complete problem.

(Partition Problem) Given an integer $T \in \mathbb{Z}_+$, consider n positive numbers $\{\alpha_i\}_{i\in[T]} \subseteq \mathbb{Z}_{++}$ having an even sum, is there a partition S_1, S_2 such that $\sum_{i\in S_1} \alpha_i = \sum_{i\in S_2} \alpha_i = \beta, S_1 \cap S_2 = \emptyset, S_1 \cup S_2 = [T]$?

We show that finding a feasible solution for a special case of DrFRAM (5) can be reduced to the partition problem. First of all, suppose there is only $N=1$ empirical sample and the Wasserstein radius $\theta=0$. We also assume that there are $K=2$ types of buses, there are $\eta_k=\beta$ buses for each $k\in[K]$ and the reduced bus capacity is $\lfloor \delta_k c_k \rfloor = C > \max_{i\in[T]} a_i$. Suppose that there are $I=T$ routes and for each route $i\in[I]$, there is only $J_i=1$ stop and the empirical alight rate $\bar{a}_{i1}=1$ and arrival rate $\bar{\lambda}_{i1}=a_i$. Under this setting, DrFRAM (5) can be reduced to the following optimization problem:

$$
\begin{aligned}
v^* = \min_{\boldsymbol{n},\boldsymbol{x}} \max_{i\in[T]} \frac{a_i}{C\sum_{k\in[2]} n_{ik}}, & \qquad (33a)\\
\text{s.t.} \quad \sum_{k\in[2]} x_{ik} = 1, \forall i\in[T], & \qquad (33b)\\
n_{ik} \le \beta x_{ik}, \forall i\in[T], \forall k\in[2], & \qquad (33c)\\
\sum_{i\in[T]} n_{ik} \le \beta, \forall k\in[2], & \qquad (33d)\\
\sum_{k\in[2]} n_{ik} \ge 1, \forall i\in[T], & \qquad (33e)\\
x_{ik}\in\{0,1\}, n_{ik}\in\mathbb{Z}^+, \forall i\in[T], \forall k\in[2]. & \qquad (33f)
\end{aligned}
$$

Thus, in the model (33), the optimal value $v^* \le \frac{1}{C}$ if and only if there exists a feasible solution $(\boldsymbol{n},\boldsymbol{x})$ satisfying the $\sum_{k\in[2]} n_{ik} \ge a_i$ for all $i\in[I]$. Or equivalently, the following set is nonempty

$$
\mathcal{T} := \{(\boldsymbol{n},\boldsymbol{x}) : n_{ik} \ge a_i x_{ik}, \forall i\in[T], \forall k\in[2], (33b)-(33f)\}.
$$

Note that $2\beta = \sum_{i\in[T]} a_i$. Thus, in set $\mathcal{T}$, we must have $n_{ik} = a_i x_{ik}$. Hence, set $\mathcal{T}$ is nonempty if and only if the following integer program is nonempty:

$$
\sum_{i\in[T]} a_{ik} x_{ik} = \beta, \sum_{k\in[2]} x_{ik} = 1, x_{ik}\in\{0,1\}, \forall i\in[T], \forall k\in[2],
$$

which is exactly equivalent to the partition problem.

Hence, checking the special case (33) having a feasible solution with its objective value no larger than $1/C$ is equivalent to solving the partition problem. This proves the NP-hardness of DrFRAM (5). □

A.2. Proof of Proposition 4

Proposition 4 *Suppose that the weight $\omega = 0$ (i.e., the passenger abandon rate for each bus stop is negligible), and there exists a matrix $\boldsymbol{\mu} = (\boldsymbol{\mu}^a, \boldsymbol{\mu}^\lambda) \in \Xi$ and a positive integer $\widehat{\alpha} \in \mathbb{Z}_{++}$ such that*

$$\widehat{a}_{ij}^{\ell} = \mu_{ij}^a, \mu_{ij}^\lambda \leq |\widehat{\lambda}_{ij}^{\ell}| \leq \widehat{\alpha}\mu_{ij}^\lambda, \forall i \in [I], \forall j \in [J_i].$$

Then the following approximation ratio holds for the naive scenario decomposition lower bound (without any no good cut or objective cut)

$$\frac{1}{\widehat{\alpha}} v^* \leq \frac{1}{N} \sum_{\ell \in [N]} v^\ell \leq v^*.$$

Proof: We split the proof into three steps.

Step 1. The second inequality holds true since we drop the nonanticipactivity constraints when computing the naive scenario decomposition lower bound.

Step 2. Next, to prove the first inequality, let us use v^D to denote the optimal value of the following nominal problem as

$$v^D(\alpha) = \min_{\boldsymbol{n},\boldsymbol{x},\boldsymbol{u}} \{\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}, \alpha\boldsymbol{\mu}) : (1b) - (1f), (7a), (7b)\} \tag{34}$$

for some positive integer $\alpha \in \mathbb{Z}_{++}$. In model (34), the function $\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}, \alpha\boldsymbol{\mu})$ is defined in (12b) by letting $\widehat{\boldsymbol{\xi}}^\ell = \alpha\boldsymbol{\mu}$.

According to the monotonicity results in Corollary 1, we have

$$\mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}, \widehat{\boldsymbol{\xi}}^\ell) \leq \mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}, \widehat{\alpha}\boldsymbol{\mu})$$

for all $\ell \in [N]$ and any feasible $(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u})$. Thus, aggregating the above inequalities for all $\ell \in [N]$, we have

$$\frac{1}{N} \sum_{\ell \in [N} \mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}, \widehat{\boldsymbol{\xi}}^\ell) \leq \mathcal{Q}(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u}, \widehat{\alpha}\boldsymbol{\mu})$$

for any feasible $(\boldsymbol{n}, \boldsymbol{x}, \boldsymbol{u})$. This implies that $v^* \leq v^D(\widehat{\alpha})$.

Similarly, we also have $v^\ell \geq v^D(1)$ for each $\ell \in [N]$. That is, $\frac{1}{N}\sum_{\ell \in [N]} v^\ell \geq v^D(1)$. Now it remains to show that $v^D(\widehat{\alpha}) \leq \widehat{\alpha} v^D(1)$.

Step 3. We observe that

Claim 1 *For any real number $q \in \mathbb{R}_+$ and positive integer number $\alpha \in \mathbb{Z}_{++}$, the following inequality must hold $\alpha\lceil q \rceil \geq \lceil \alpha q \rceil$.*

Proof: This is simply because $\alpha\lceil q \rceil \geq \alpha q$ and the former is an integer. ◇

Suppose that $(\boldsymbol{n}^*, \boldsymbol{x}^*, \boldsymbol{u}^*)$ is an optimal first-stage solution and $(\boldsymbol{L}^1, \bar{\boldsymbol{L}}^1, \boldsymbol{w}^1, \boldsymbol{y}^1, \boldsymbol{E}^1)$ is an optimal second-stage decision to model (34) with $\alpha = 1$.

Now let us define

$$
\begin{aligned}
&L_{ij}^{\widehat{\alpha}} = \min\left\{\sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor, \widehat{\alpha} L_{ij}^1\right\}, \forall i \in [I], \forall j \in [J_i],\\
&\bar{L}_{ij-1}^{\widehat{\alpha}} = \lceil (1-\mu_{ij}^a) L_{i,j-1}^{\widehat{\alpha}} \rceil, \forall i \in [I], \forall j \in [J_i],\\
&y_{ij}^{\widehat{\alpha}} = \mathbb{I}\left(L_{ij}^{\widehat{\alpha}} < \sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor\right), \forall i \in [I], \forall j \in [J_i],\\
&E_1^{\widehat{\alpha}} = \max_{i\in[I], j\in[J_i]} \frac{L_{ij}^{\widehat{\alpha}}}{\sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor},\\
&E_2^{\widehat{\alpha}} = \max_{i\in[I], j\in[J_i]} (\widehat{\alpha}\mu_{ij}^{\lambda})^{-1} \left(\bar{L}_{i,j-1}^{\widehat{\alpha}} + \widehat{\alpha}\mu_{ij}^{\lambda} - \sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor\right)_+,\\
&w_{ikr}^{\widehat{\alpha}} = E_1^{\widehat{\alpha}} u_{ikr}^*, \forall i \in [I], \forall k \in [K], \forall r \in [R_k].
\end{aligned}
$$

According to the definition of the model (34) with $\alpha = \widehat{\alpha}$, we see that $(\boldsymbol{n}^*, \boldsymbol{x}^*, \boldsymbol{u}^*)$ is a feasible first-stage solution and $(\boldsymbol{L}^{\widehat{\alpha}}, \bar{\boldsymbol{L}}^{\widehat{\alpha}}, \boldsymbol{w}^{\widehat{\alpha}}, \boldsymbol{y}^{\widehat{\alpha}}, \boldsymbol{E}^{\widehat{\alpha}})$ satisfies constraints (2c), (9), (11b)–(11d). It remains to show that constraints (10) also hold.

Before proceeding, we observe the following fact:

Claim 2 *For any $i \in [I], j \in [J_i]$, we must have $\bar{L}_{ij-1}^{\widehat{\alpha}} \le \widehat{\alpha}\bar{L}_{ij-1}^1$.*

Proof: Note that

$$
\bar{L}_{ij-1}^{\widehat{\alpha}} = \lceil (1-\mu_{ij}^a) L_{i,j-1}^{\widehat{\alpha}} \rceil \le \lceil (1-\mu_{ij}^a) \widehat{\alpha} L_{i,j-1}^1 \rceil \le \widehat{\alpha} \lceil (1-\mu_{ij}^a) L_{i,j-1}^1 \rceil := \widehat{\alpha}\bar{L}_{ij-1}^1,
$$

where the first inequality is due to the definition of $L_{i,j-1}^{\widehat{\alpha}}$ and the second one is due to Claim 1. $\diamond$
There are two cases:

- When $L_{ij}^{\widehat{\alpha}} < \sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor$, it is sufficient to show that $L_{ij}^{\widehat{\alpha}} \ge \bar{L}_{ij-1}^{\widehat{\alpha}} + \widehat{\alpha}\mu_{ij}^{\lambda}$. This must be true since we have

  $$
  L_{ij}^{\widehat{\alpha}} = \widehat{\alpha} L_{ij}^1, \bar{L}_{ij-1}^{\widehat{\alpha}} \le \widehat{\alpha}\bar{L}_{ij-1}^1, L_{ij}^1 \ge \bar{L}_{ij-1}^1 + \mu_{ij}^{\lambda}
  $$

  where the first inequality is due to Claim 2 and the second one is due to the fact that $(\boldsymbol{L}^1, \bar{\boldsymbol{L}}^1, \boldsymbol{w}^1, \boldsymbol{y}^1, \boldsymbol{E}^1)$ is an optimal (and of course feasible) second-stage decision to the model (34) with $\alpha = 1$.

- When $L_{ij}^{\widehat{\alpha}} = \sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor$, we have $y_{ij}^{\alpha} = 0$. It is sufficient to show that $L_{ij}^{\widehat{\alpha}} \geq \bar{L}_{ij-1}^{\widehat{\alpha}}$. This is true since

$$\bar{L}_{ij-1}^{\widehat{\alpha}} = \lceil (1-\mu_{ij}^a) L_{i,j-1}^{\widehat{\alpha}} \rceil \leq \sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor := L_{ij}^{\widehat{\alpha}}$$

  where the inequality is due to the fact that $L_{i,j-1}^{\widehat{\alpha}} \leq \sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor$.

Q.E.D.

Finally, we observe that the objective function is

$$\begin{aligned} E_1^{\widehat{\alpha}} + \omega E_2^{\widehat{\alpha}} &= \max_{i\in[I], j\in[J_i]} \frac{L_{ij}^{\widehat{\alpha}}}{\sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor} + \omega \max_{i\in[I], j\in[J_i]} (\widehat{\alpha}\mu_{ij}^{\lambda})^{-1} \left(\bar{L}_{i,j-1}^{\widehat{\alpha}} + \widehat{\alpha}\mu_{ij}^{\lambda} - \sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor \right)_+ \\ &= \max_{i\in[I], j\in[J_i]} \frac{\min\left\{ \sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor, \widehat{\alpha} L_{ij}^1 \right\}}{\sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor} \\ &\leq \max_{i\in[I], j\in[J_i]} \frac{\widehat{\alpha}\min\left\{ \sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor, L_{ij}^1 \right\}}{\sum_{k\in[K]} n_{ik}^* \lfloor \delta_k c_k \rfloor} = \widehat{\alpha} E_1^1 := \widehat{\alpha} v^D(1) \end{aligned}$$

where the inequality is due to the fact that $\min\{a, \widehat{\alpha} b\} \leq \widehat{\alpha} \min\{a, b\}$ for any non-negative numbers a, b and positive integer $\widehat{\alpha}$.

Since $(\boldsymbol{n}^*, \boldsymbol{x}^*, \boldsymbol{u}^*)$ is a feasible first-stage solution and $(\boldsymbol{L}^{\widehat{\alpha}}, \bar{\boldsymbol{L}}^{\widehat{\alpha}}, \boldsymbol{w}^{\widehat{\alpha}}, \boldsymbol{y}^{\widehat{\alpha}}, \boldsymbol{E}^{\widehat{\alpha}})$ is a feasible second-stage solution to model (34) with the objective value at most $\widehat{\alpha} v^D(1)$, this proves that $v^D(\widehat{\alpha}) \leq \widehat{\alpha} v^D(1)$. This completes the proof. □

A.3. Proof of Lemma 2

Lemma 2 *The following characterization holds for* $\mathrm{conv}(X_{ij\ell k}^2 \wedge \{x_{ik} = 1\})$*, i.e.,*

$$\mathrm{conv}(X_{ij\ell k}^2 \wedge \{x_{ik} = 1\}) = \mathrm{conv}(\bar{X}_{ij\ell k}^2) \bigcap \left\{ (\boldsymbol{w}_{i::}^{\ell}, L_{ij}^{\ell}) : \sum_{r\in[R_k]} 2^{r-1} \lfloor \delta_k c_k \rfloor w_{ikr}^{\ell} \geq L_{ij}^{\ell} \right\},$$

where

$$\bar{X}_{ij\ell k}^2 = \left\{ (\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}, L_{ij}^{\ell}, \boldsymbol{w}_{i::}^{\ell}, E_1^{\ell}) : (\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}) \in X_{ik}^1, w_{ikr}^{\ell} = u_{ikr} E_1^{\ell}, L_{ij}^{\ell} \in \mathbb{R}_+, E_1^{\ell} \in [0,1], x_{ik} = 1 \right\}.$$

Proof: Since the convex hull of the intersection of two sets is contained in the intersection of the convex hulls of two sets, we must have

$$\mathrm{conv}(X_{ij\ell k}^2 \wedge \{x_{ik} = 1\}) \subseteq \mathrm{conv}(\bar{X}_{ij\ell k}^2) \bigcap \left\{ (\boldsymbol{w}_{i::}^{\ell}, L_{ij}^{\ell}) : \sum_{r\in[R_k]} 2^{r-1} \lfloor \delta_k c_k \rfloor w_{ikr}^{\ell} \geq L_{ij}^{\ell} \right\}.$$

It remains to show that

$$\mathrm{conv}(X_{ij\ell k}^2 \wedge \{x_{ik} = 1\}) \supseteq \mathrm{conv}(\bar{X}_{ij\ell k}^2) \bigcap \left\{ (\boldsymbol{w}_{i::}^{\ell}, L_{ij}^{\ell}) : \sum_{r\in[R_k]} 2^{r-1} \lfloor \delta_k c_k \rfloor w_{ikr}^{\ell} \geq L_{ij}^{\ell} \right\}.$$

Indeed, for any $(\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}, L^{\ell}_{ij}, \boldsymbol{w}^{\ell}_{i::}, E^{\ell}_1) \in \text{conv}(\bar{X}^2_{ij\ell k}) \bigcap \left\{(\boldsymbol{w}^{\ell}_{i::}, L^{\ell}_{ij}) : \sum_{r\in[R_k]} 2^{r-1}\lfloor \delta_k c_k \rfloor w^{\ell}_{ikr} \geq L^{\ell}_{ij}\right\}$. There exists a finite collection $\{(\boldsymbol{x}^{\tau}_{i:}, \boldsymbol{u}^{\tau}_{i::}, L^{\ell\tau}_{ij}, \boldsymbol{w}^{\ell\tau}_{i::}, E^{\ell\tau}_1)\}_{\tau\in[q]}$ and $\{\alpha_\tau\}_{\tau\in[q]} \subset [0,1]$ such that

$$\sum_{\tau\in[q]} \alpha_\tau(\boldsymbol{x}^{\tau}_{i:}, \boldsymbol{u}^{\tau}_{i::}, L^{\ell\tau}_{ij}, \boldsymbol{w}^{\ell\tau}_{i::}, E^{\ell\tau}_1) = (\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}, L^{\ell}_{ij}, \boldsymbol{w}^{\ell}_{i::}, E^{\ell}_1), \sum_{\tau\in[q]} \alpha_\tau = 1$$

and $\beta := \sum_{\tau\in[q]} \alpha_\tau \sum_{r\in[R_k]} 2^{r-1}\lfloor \delta_k c_k \rfloor w^{\ell\tau}_{ikr} \geq L^{\ell}_{ij}$.

Now let us define

$$\widehat{L}^{\ell\tau}_{ij} = \frac{L^{\ell}_{ij}}{\beta} \sum_{r\in[R_k]} 2^{r-1}\lfloor \delta_k c_k \rfloor w^{\ell\tau}_{ikr}$$

for each $\tau \in [q]$. Clearly, we have

$$\sum_{\tau\in[q]} \alpha_\tau(\boldsymbol{x}^{\tau}_{i:}, \boldsymbol{u}^{\tau}_{i::}, \widehat{L}^{\ell\tau}_{ij}, \boldsymbol{w}^{\ell\tau}_{i::}, E^{\ell\tau}_1) = (\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}, L^{\ell}_{ij}, \boldsymbol{w}^{\ell}_{i::}, E^{\ell}_1),$$
$$(\boldsymbol{x}^{\tau}_{i:}, \boldsymbol{u}^{\tau}_{i::}, \widehat{L}^{\ell\tau}_{ij}, \boldsymbol{w}^{\ell\tau}_{i::}, E^{\ell\tau}_1) \in X^2_{ij\ell k} \wedge \{x_{ik} = 1\}, \forall \tau \in [q].$$

Thus, $(\boldsymbol{x}_{i:}, \boldsymbol{u}_{i::}, L^{\ell}_{ij}, \boldsymbol{w}^{\ell}_{i::}, E^{\ell}_1) \in \text{conv}(X^2_{ij\ell k} \wedge \{x_{ik} = 1\})$. □

Appendix B: Testing Integrality of L_{ij}

We tested six small instances to compare the objective value of MILP formulations with the integer $\boldsymbol{L}$ as well as the objective value of MILP with relaxing the integrality of $\boldsymbol{L}$, where the number of bus routes I is 5 and the number of scenarios N is 6. There are $K = 3$ types of buses with nominal capacities $c_1 = 60, c_2 = 80, c_3 = 120$. Each route has $10 \sim 40$ stops. We let $\boldsymbol{\eta} = (4, 6, 3), \delta = 0.25, \omega = 5, \theta = 0, \gamma = 100$. The passenger arrival rates were assumed to follow uniform distributions with the minimum value ranging from 2 to 20 and the maximum one from 22 to 40, and the proportion of passengers alighting from the bus was randomly generated from triangular distributions with the lower limit ranging from 0 to 0.2, the upper limit range from 0.6 to 1, and the mode ranging from 0.45 to 0.55. The results are shown in Table 11. By relaxing the integrality of $\boldsymbol{L}$, the solution time decreases around 15%, which is not very significant since our model is for strategic planning purposes. However, we see that the objective value by relaxing passenger number to be continuous can decrease by around 14%, which can be quite significant and cause misleading decisions. Relaxing the integrality of $\boldsymbol{L}$ can result in a misleading solution as shown in Table 12, where different routes have different bus assignments. Dominant routes in each scenario (i.e., the ones having the largest weighted sum of bus utilization and passenger abandon rates) are also different. Since there is no big improvement in the solution time, we believe that keeping integer $\boldsymbol{L}$ is necessary for the sake of solution quality.

Table 11 Objective Value of MILP with Integer L and MILP with Continuous L

Instance	Integer $\boldsymbol{L}$		Continuous $\boldsymbol{L}$		Relative
	Obj.Val	Time	Obj.Val	Time	Difference(%)
1	1.148	11.0	0.948	10.3	17.4
2	1.241	16.4	1.034	13.5	16.7
3	1.178	22.9	1.022	20.9	13.3
4	1.106	22.5	0.986	17.1	10.8
5	1.104	18.8	0.978	14.2	11.5
6	1.330	13.3	1.163	12.0	12.5

Table 12 Bus assignment of MILP with Integer L and MILP with Continuous L

	Integer $\boldsymbol{L}$				Continuous $\boldsymbol{L}$			
Obj.Val	0.942				0.916			
Route	Dominant Scenario	Bus 1	Bus 2	Bus 3	Dominant Scenario	Bus 1	Bus 2	Bus 3
1	2			1				1
2				2	1-6		4	
3	1, 3-6		3				2	
4			3					2
5		3				4		