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

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, Höcher 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., \( \mathbf{x}, \mathbf{A} \)) to denote vectors and matrices and use corresponding non-bold letters to denote their components. Given a vector or matrix \( \mathbf{x} \), its zero norm \( \|\mathbf{x}\|_0 \) 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 \( i \)th 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 \( \tilde{\xi} \) denote a random vector and denote its realizations by \( \xi \). We use superscript \( k \in [N] \) to denote the index of scenario \( k \). For a matrix \( \mathbf{A} \), we let \( \mathbf{A}_i \) denote \( i \)th row of \( \mathbf{A} \) and \( \mathbf{A}_j \) denote \( j \)th column of \( \mathbf{A} \). 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 Shnaidman (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 $\eta_k$ 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 $\tilde{a}, \tilde{\lambda}$ are both random since they can vary over time and change with severeness of the pandemic. For notational convenience, we let $\tilde{\xi} = (\tilde{a}, \tilde{\lambda})$. 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 $\xi$ of the random parameters $\tilde{\xi}$, we let $L_{ij}(\xi)$ denote the number of passengers remaining on a bus right after the $j$th 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 $L_{ij}(\xi)$ 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:

$$(\text{DrFRAM}) \quad v^* = \min \sup_{n, x, \tilde{\xi} \in \mathcal{P}} \mathbb{E}_p \left[ Q \left( n, x, \tilde{\xi} \right) \right], \quad (1a)$$

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

$n_{ik} \leq \eta_k x_{ik}, \forall i \in [I], \forall k \in [K]$, \quad (1c)

$^1$ Our numerical result in Appendix B shows that relaxing the integrality of $L_{ij}(\xi)$ can cause a much smaller objective value and may even lead to a misleading conclusion.
\[ \sum_{i \in [I]} n_{ik} \leq \eta_k, \forall k \in [K], \quad (1d) \]
\[ \sum_{k \in [K]} n_{ik} \geq 1, \forall i \in [I], \quad (1e) \]
\[ x_{ik} \in \{0, 1\}, n_{ik} \in \mathbb{Z}^+, \forall i \in [I], \forall k \in [K]. \quad (1f) \]

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

\[
Q(n, x, \xi) = \min_{L(\xi)} Q(n, x, \xi, L(\xi)) := \left\{ \max_{i \in [I], j \in [J_i]} \frac{L_{ij}(\xi)}{\sum_{k \in [K]} n_{ik} \lfloor \delta_k c_k \rfloor} \right\} + \omega \max_{i \in [I], j \in [J_i]} \max \left[ 0, 1 + \frac{1}{\lambda_{ij}} \left[ (1 - a_{ij}) L_{i,j-1}(\xi) - \sum_{k \in [K]} n_{ik} \lfloor \delta_k c_k \rfloor \right] \right], \quad (2a)
\]

s.t. \( L_{ij}(\xi) = \min \left\{ \sum_{k \in [K]} n_{ik} \lfloor \delta_k c_k \rfloor, \lceil (1 - a_{ij}) L_{i,j-1}(\xi) + \lambda_{ij} \rceil \right\}, \forall i \in [I], \forall j \in [J_i], \quad (2b) \]
\[ L_{i0}(\xi) = 0, L_{ij}(\xi) \in \mathbb{Z}^+, \forall i \in [I], \forall j \in [J_i], \quad (2c) \]

where \( \omega \geq 0 \) 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 \( L(\xi) \).

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 \( \tilde{\mathbf{\xi}} \) constructed using i.i.d. historical data \( \mathbf{Z} = \{ \bar{\mathbf{\xi}}^\ell \}_{\ell \in [N]} \) such that \( \tilde{\mathbf{\xi}} \{ \tilde{\mathbf{\xi}} = \bar{\mathbf{\xi}}^\ell \} = 1/N \), 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:

\[
\mathcal{P}_q = \{ \mathbb{P} : W_q (\mathbb{P}, \tilde{\mathbb{P}}_{\mathbf{\xi}}) \leq \theta \}, \tag{3}
\]

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

\[
W_q (\mathbb{P}_1, \mathbb{P}_2) = \inf \left\{ \sqrt{\frac{1}{Q} \int_{\Xi \times \Xi} ||\mathbf{\xi}_1 - \mathbf{\xi}_2||^q Q (d\mathbf{\xi}_1, d\mathbf{\xi}_2) : Q \text{ is a joint distribution of } \tilde{\mathbf{\xi}}_1 \text{ and } \tilde{\mathbf{\xi}}_2 \text{ with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively} \right\},
\]

where the support

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

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

\[
W_\infty (\mathbb{P}_1, \mathbb{P}_2) = \inf \left\{ Q - \text{ess sup} ||\tilde{\mathbf{\xi}}_1 - \tilde{\mathbf{\xi}}_2|| : Q \text{ is a joint distribution of } \tilde{\mathbf{\xi}}_1 \text{ and } \tilde{\mathbf{\xi}}_2 \text{ with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively} \right\}.
\]

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

\[
Q - \text{ess sup} ||\tilde{\mathbf{\xi}}_1 - \tilde{\mathbf{\xi}}_2|| = \inf \left\{ \Delta : Q(\{ ||\tilde{\mathbf{\xi}}_1 - \tilde{\mathbf{\xi}}_2|| > \Delta \} = 0 \right\}.
\]

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

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

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 \( \theta \) amount from one of the empirical data \( Z = \{ \bar{\mathbf{\xi}}^k \}_{k \in [N]} \subset \Xi. \)
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-$\infty$ Wasserstein ambiguity set $\mathcal{P} = \mathcal{P}_\infty$, DrFRAM (1) can be equivalently represented as the following deterministic counterpart:

\begin{equation}
(\text{DrFRAM}) \quad v^* = \min_{n, x} \left\{ \frac{1}{N} \sum_{\ell \in [N]} \sup_{\xi^{\ell} \in \Xi, \|\xi^{\ell} - \zeta^{\ell}\| \leq \theta} Q(n, x, \xi^{\ell}) : (1b) - (1f) \right\}.
\end{equation}

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 $Q(n, x, \xi)$ and the second-stage objective function $Q(n, x, \xi, L(\xi))$. 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 $Q(n, x, \xi, L(\xi))$. 

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

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

Corollary 1  For the recourse function \( Q(n, x, \xi) \), we have the following sensitivity results:

- Recourse function \( Q(n, x, \xi) \) is monotone non-increasing as the first-stage decision \( n_{ik} \) increases for some \( i \in [I] \) and \( k \in [K] \);
- Recourse function \( Q(n, x, \xi) \) is monotone non-decreasing as the passenger arrival rate \( \lambda_{ij} \) increases for some \( i \in [I] \) and \( j \in [J_i] \); and
- Recourse function \( Q(n, x, \xi) \) 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-\( \infty \) Wasserstein ambiguity set. Specifically, since random parameters \( \tilde{a} \) and \( \tilde{\lambda} \) have different magnitude, we choose the weighted \( \ell_\infty \) norm as \( \| \xi \| = \max\{\gamma \| a \|_\infty, \| \lambda \|_\infty\} \) (recall that we let \( \xi = (a, \lambda) \)), where the positive weight \( \gamma > 0 \) is to demonstrate the balance of the importance of both parameters. In
practice, one can choose a proper $\gamma$ such that the scaled vector $\gamma a$ and vector $\lambda$ have a similar order of magnitude.

In the spirits of Corollary 1 and the weighted $\ell_\infty$ norm, DrFRAM (5) admits the following equivalent representation.

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

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

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

**Proof:** The simplification of the inner supremum in DrFRAM (5) follows the monotonicity results in Corollary 1 and the definition of the weighted $\ell_\infty$ norm. \hfill \Box

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.,

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

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

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

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

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 $\xi = \hat{\xi}^\ell$ for each $\ell \in [N]$. First, to suppress the notation, we let $L^\ell := L(\hat{\xi}^\ell)$.
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

\[ L_{ij}^\ell \geq \min \left\{ \sum_{k \in [K]} n_{ik} [\delta_k c_k], [(1 - a_{ij}^\ell)L_{i,j-1}^\ell + \lambda_{ij}], \forall i \in [I], \forall j \in [J] \right\}, \]

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

\[ \bar{L}_{i,j-1}^\ell \geq (1 - a_{ij}^\ell)L_{i,j-1}^\ell, \bar{L}_{i,j-1}^\ell \in \mathbb{Z}_+, \forall i \in [I], \forall j \in [J]. \] (8)

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

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

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

\[ L_{ij}^\ell \geq \sum_{k \in [K]} n_{ik} [\delta_k c_k] - M y_{ij}, L_{ij}^\ell \geq \bar{L}_{i,j-1}^\ell + \bar{\lambda}_{ij}^\ell - M(1 - y_{ij}), y_{ij} \in \{0, 1\}, \forall i \in [I], \forall j \in [J], \]

\[ L_{ij}^\ell \geq \sum_{k \in [K]} n_{ik} [\delta_k c_k] - (M - \bar{\lambda}_{ij}^\ell)y_{ij}, L_{ij}^\ell \geq \bar{L}_{i,j-1}^\ell + \bar{\lambda}_{ij}^\ell, y_{ij} \in \{0, 1\}, \forall i \in [I], \forall j \in [J], \] (10)

where we choose \(M = \max_{k \in [K]} \eta_k [\delta_k c_k]\) 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 \(\bar{L}_{i,j-1} \leq L_{i,j-1}^\ell\).

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

For each \(\ell \in [N]\), let us use \(E_1^\ell, E_2^\ell\) 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

\[ Q(n, x, \bar{\xi}^\ell, \bar{L}^\ell) = E_1^\ell + \omega E_2^\ell \]

where

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

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

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

\[ \sum_{k \in [K]} \sum_{r \in [R_k]} 2r^{-1} [\delta_k c_k] E_1^\ell u_{ikr} \geq L_{ij}^\ell, \forall i \in [I], \forall j \in [J], E_1^\ell \in [0, 1], \]
Introducing McCormick inequalities to linearize the bilinear terms, we have

\[
\sum_{k \in [K]} \sum_{r \in [R_k]} 2^{r-1} \left[ \delta_k c_k \right] w_{ikr}^\ell \geq L_{ij}^\ell, \forall i \in [I], \forall j \in [J], E_1^\ell \in [0, 1], \quad (11c)
\]

\[
w_{ikr}^\ell \geq 0, w_{ikr}^\ell \geq E_1^\ell + u_{ikr} - 1, w_{ikr}^\ell \leq u_{ikr}, \forall k \in [K], \forall i \in [I], \forall r \in [R]. \quad (11d)
\]

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:

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

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

\[
Q(n, x, u, \hat{\xi}^\ell) = \min_{L^\ell, \bar{L}^\ell, w^\ell, y^\ell, E^\ell} \left\{ E_1^\ell + \omega E_2^\ell : (2c), (9), (10), (11b) - (11d) \right\}. \quad (12b)
\]

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

\[
v^\ell := \min_{n, x, u} \left\{ Q \left( n, x, u, \hat{\xi}^\ell \right) : (1b) - (1f), (7a), (7b) \right\}. \quad (13)
\]

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 \(\ell\), given a first-stage binary solution \((\bar{x}^\ell, \bar{u}^\ell)\) (we do not need to include \(\bar{n}^\ell\) since it can be represented by \(\bar{u}^\ell\)), we add the following no-good cut

\[
\sum_{i \in [I], k \in [K]} \left( 1 - \bar{x}_{ik}^\ell \right) x_{ik} + \sum_{r \in [R_k]} \left( 1 - \bar{u}_{ikr}^\ell \right) u_{ikr} \geq 1
\]
into the $\ell$th 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 $G_\nu \subset [N], \nu \in [N]$ 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 $N$ subproblems, i.e., for each $\nu \in [N]$, we solve

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

The detailed implementation can be found in Algorithm 2.

---

**Algorithm 1: Scenario Decomposition for Solving DrFRAM**

1. Initialization: Set $\tau = 0, \epsilon > 0$, $\mathcal{X}' = \{(n, x, u) : (1b) - (1f), (7a), (7b)\}$ for all $\ell \in [N]$, $LB = -\infty, UB = \infty$;
2. while $UB - LB > \epsilon$ do
   3. for $\ell \in [N]$ do
      4. $$(\bar{n}^\ell, \bar{x}^\ell, \bar{u}^\ell) \in \arg\min \{Q(n, x, u, \hat{\xi}^\ell) : (n, x, u) \in \mathcal{X}'\};$$
      5. $LB = \max \{LB, \frac{1}{N} \sum_{\ell \in [N]} Q(n, x, u, \hat{\xi}^\ell)\};$
      6. $UB = \min \{UB, \frac{1}{N} \sum_{\ell \in [N]} Q(n, x, u, \hat{\xi}^\ell)\};$
      7. $\mathcal{X}' \to \mathcal{X}' \setminus \{(\bar{n}^\ell, \bar{x}^\ell, \bar{u}^\ell)\}$ for each $\ell \in [N];$
      8. $\tau \leftarrow \tau + 1$

---

**Algorithm 2: Scenario Decomposition with Grouping for Solving DrFRAM**

1. Initialization: Set $\tau = 0, \epsilon > 0$, integer $\mathcal{N}$ (divisible of $N$), cluster $G_\nu$, and set $\mathcal{X}'' = \bigcap_{\ell \in G_\nu} \{(n, x, u) : (1b) - (1f), (7a), (7b)\}$ for all $\nu \in [N], LB = -\infty, UB = \infty$;
2. while $UB - LB > \epsilon$ do
   3. for $\nu \in [N]$ do
      4. $$(\bar{n}''\nu, \bar{x}''\nu, \bar{u}''\nu) \in \arg\min \left\{ v'' = \frac{1}{|G_\nu|} \sum_{\ell \in G_\nu} Q(n, x, u, \hat{\xi}^\ell) : (n, x, u) \in \mathcal{X}''\nu \right\};$$
      5. $LB = \max \{LB, \frac{1}{N} \sum_{\nu \in [N]} v''\nu\};$
      6. $UB = \min \{UB, \frac{1}{N} \sum_{\nu \in [N]} \frac{1}{N} \sum_{\ell \in [N]} Q(n, x, u, \hat{\xi}^\ell)\};$
      7. $\mathcal{X}'' \to \mathcal{X}'' \setminus \{(\bar{n}''\nu, \bar{x}''\nu, \bar{u}''\nu)\}$ for each $\ell \in [N];$
      8. $\tau \leftarrow \tau + 1$
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
\[ E_1^\ell + \omega E_2^\ell \geq Q(\bar{n}^\ell, \bar{x}^\ell, \bar{u}^\ell) + \tilde{c}^\ell \] into the $\ell$th subproblem, where $(\bar{n}^\ell, \bar{x}^\ell, \bar{u}^\ell)$ denotes an optimal solution of $\ell$th subproblem and $\tilde{c}^\ell > 0$. 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**

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

Finally, we conclude this section by showing that if the data from distributionally robust counterparts (i.e., $\{\tilde{\xi}^\ell\}_{\ell \in [N]}$) 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 $v^\ast$).

**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 $\mu = (\mu^a, \mu^\lambda) \in \Xi$ and a positive integer $\hat{\alpha} \in \mathbb{Z}^+$ such that
\[ \hat{a}_{ij}^\ell = \mu_{ij}^a, \mu_{ij}^\lambda \leq |\hat{\lambda}_{ij}^\ell| \leq \hat{\alpha} \mu_{ij}^\lambda, \forall i \in [I], \forall j \in [J_i]. \]

Then the following approximation ratio holds for the naive scenario decomposition lower bound (without any no good cut or objective cut)
\[ \frac{1}{\hat{\alpha}} v^\ast \leq \frac{1}{N} \sum_{\ell \in [N]} v^\ell \leq v^\ast. \]

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 $(x_i, u_i)$ by projecting out variables $n_i$. That is, for each $i \in [I]$, let us define

$$X^1_i = \{ (x_i, u_i) : \exists n_i, (1b) - (1c), (7a), (7b) \}. \quad \text{(16a)}$$

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

$$X^2_{ij\ell} = \{ (x_i, u_i, L_{ij}) \in X^1_i \times \mathbb{R}_+ : (11c), (11d) \}. \quad \text{(16b)}$$

5.1. Convexification of Set $X^1_i$

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

$$X^1_i = \left( \bigvee_{k \in [K]} (X^{1}_{ik} \land \{ x_{ik} = 1 \}) \right) \land \left\{ x_i \in \{0, 1\}^K : \sum_{k \in [K]} x_{ik} = 1 \right\}$$

where

$$X^{1}_{ik} := \left\{ (x_i, u_i) : \sum_{r \in [R_k]} 2^{-1} u_{ikr} \leq \eta_k x_{ik}, u_{ikr} \in \{0, 1\}, \forall r \in [R_k] \right\}.$$ 

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

$$\text{conv} \left( X^{1}_{ik} \land \{ x_{ik} = 1 \} \right) = \left\{ (x_i, u_i) : u_{ikr} \leq \sum_{\tau \in I_r} u_{ikr}, u_{ikr} \in \{0, 1\} \right\},$$

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

Our convex hull description of set $X^1_i$ relies on the well-known disjunctive programming (Balas 1979).
Lemma 1 Suppose $\bar{s}$ polyhedra $\bar{X}^i = \{ \bar{y} \in \mathbb{R}^n : A^i \bar{y} \leq b^i \}$ with $A^i \in \mathbb{R}^{m_i \times n}$ and $b \in \mathbb{R}^{m_i}$ for each $i \in [s]$ share the same recession cone. Then the following result holds

$$
\text{conv} \left\{ \left( \bigvee_{i \in [s]} (\bar{X}^i \wedge \{ \lambda_i = 1 \}) \right) \wedge \left\{ \lambda \in \{0, 1\}^s : \sum_{i \in [s]} \lambda_i = 1 \right\} \right\}
$$

$$= \left\{ (\bar{y}, \lambda) \in \mathbb{R}^n \times \{0, 1\}^s : A^i \bar{y}^i \leq b^i \lambda_i, \forall i \in [s] ; \sum_{i \in [s]} \bar{y}^i = \bar{y} \right\}.
$$

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 $X^1_i$. This result is summarized below.

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

$$\text{conv} \left( X^1_i \right) = \left\{ \begin{array}{l}
\mathbf{x}_i \in [0, 1]^K, \sum_{k \in [K]} x_{ik} = 1, \\
(\mathbf{x}_i, \mathbf{u}_{ikr}) : u_{ikr} + \sum_{r \in I} u_{ikr} \leq |I_r| x_{ik}, \forall r \in [I_k] \setminus I, \forall k \in [K], \\
0 \leq u_{ikr} \leq x_{ik}, \forall r \in [I_k], \forall k \in [K]
\end{array} \right\}. \quad (17)
$$

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

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

$$X^2_{ij\ell} = \left( \bigvee_{k \in [K]} (X^2_{ij\ell k} \wedge \{ x_{ik} = 1 \}) \right) \wedge \left\{ \mathbf{x}_i \in \{0, 1\}^K : \sum_{k \in [K]} x_{ik} = 1 \right\}, \quad (18)
$$

where

$$X^2_{ij\ell k} \wedge \{ x_{ik} = 1 \} := \left\{ (\mathbf{x}_i, \mathbf{u}_{ikr}, L^\ell_{ij}, \mathbf{w}^\ell_{ikr}, E^\ell_1) : (\mathbf{x}_i, \mathbf{u}_{ikr}) \in X^1_{ik}, \sum_{r \in [I_k]} 2^{r-1} |\delta_c k| w_{ikr}^\ell \geq L^\ell_{ij}, \\
\mathbf{w}^\ell_{ikr} = u_{ikr} E^\ell_1, L^\ell_{ij} \in \mathbb{R}_+, E^\ell_1 \in [0, 1], x_{ik} = 1 \right\}.
$$

We first observe that the convex hull of set $X^2_{ij\ell k} \wedge \{ x_{ik} = 1 \}$ is equal to the intersection of the constraint $\sum_{r \in [I_k]} 2^{r-1} |\delta_c k| w_{ikr}^\ell \geq L^\ell_{ij}$ and the convex hull of the feasible region subject to the remaining constraints.

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

$$\text{conv}(X^2_{ij\ell k} \wedge \{ x_{ik} = 1 \}) = \text{conv}(\bar{X}^2_{ij\ell k}) \cap \left\{ (\mathbf{w}^\ell_{ikr}, L^\ell_{ij}) : \sum_{r \in [I_k]} 2^{r-1} |\delta_c k| w_{ikr}^\ell \geq L^\ell_{ij} \right\},
$$

where

$$\bar{X}^2_{ij\ell k} = \left\{ (\mathbf{x}_i, \mathbf{u}_{ikr}, L^\ell_{ij}, \mathbf{w}^\ell_{ikr}, E^\ell_1) : (\mathbf{x}_i, \mathbf{u}_{ikr}) \in X^1_{ik}, w_{ikr} = u_{ikr} E^\ell_1, L^\ell_{ij} \in \mathbb{R}_+, E^\ell_1 \in [0, 1], x_{ik} = 1 \right\}.$$
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 \( \text{conv}(X^2_{ij\ell k} \land \{x_{ik} = 1\}) \) can be completely described using the result in Proposition 5 and the fact that variables \( x_i \) are independent of others and constraint system with respect to \( x_i \) 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 \( \text{conv}(X^2_{ij\ell k} \land \{x_{ik} = 1\}) \) admit the following complete descriptions:

- \[
\text{conv}(\bar{X}^2_{ij\ell k}) = \left\{ (x_i; u_i; L^f_{ij}, w^f_{ikr}; E^f_1) : \begin{align*}
  &x_i : [0, 1]^K, x_{ik} = 1, E^f_1 \in [0, 1], \\
  &w_{ikr} + \sum_{\tau \in L_r} w_{ik\tau} \leq |I_r| E^f_1, \forall r \in [R_k] \setminus I, \\
  &|I_r|(1 - E^f_1), \forall r \in [R_k] \setminus I, \\
  &0 \leq w_{ikr} \leq E^f_1, \forall r \in [R_k], \\
  &0 \leq u_{ikr} - w_{ikr} \leq (1 - E^f_1), \forall r \in [R_k] \\
\end{align*}\right\} ; (19)
\]

and

- \[
\text{conv}(X^2_{ij\ell k} \land \{x_{ik} = 1\}) = \left\{ (x_i; u_i; L^f_{ij}, w^f_{ikr}; E^f_1) : \begin{align*}
  &x_i : [0, 1]^K, x_{ik} = 1, E^f_1 \in [0, 1], \\
  &w_{ikr} + \sum_{\tau \in L_r} w_{ik\tau} \leq |I_r| E^f_1, \forall r \in [R_k] \setminus I, \\
  &|I_r|(1 - E^f_1), \forall r \in [R_k] \setminus I, \\
  &0 \leq w_{ikr} \leq E^f_1, \forall r \in [R_k], \\
  &0 \leq u_{ikr} - w_{ikr} \leq (1 - E^f_1), \forall r \in [R_k], \\
  &\sum_{r \in [R_k]} 2^{r-1} [\delta_k c_k] w_{ikr}^f \geq L^f_{ij}
\end{align*}\right\}. (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 \( i \in [I], j \in [J], \) and \( \ell \in [N], \) the set \( \text{conv}(X_{ij\ell}^2) \) admits the following complete description:

\[
\text{conv}(X_{ij\ell}^2) = \left\{ (x_i, u_{ij}, L_{ij\ell}, w_{ij\ell}, E_{ij\ell}^\ell) : \begin{align*}
&x_i \in [0, 1]^K, \sum_{k \in [K]} x_{ik} = 1, \sum_{k \in [K]} \hat{E}_{1ik}^\ell = E_{1ij}^\ell, \hat{E}_{1ik}^\ell \leq x_{ik}, \\
&w_{ikr} + \sum_{\tau \in \mathcal{I}_r} w_{ikr} \leq |\mathcal{I}_r| E_{1ikr}^\ell, \forall r \in [R_k] \setminus \mathcal{I}, \forall k \in [K], \\
&u_{ikr} + \sum_{\tau \in \mathcal{I}_r} u_{ikr} - \sum_{\tau \in \mathcal{I}_r} w_{ikr} \leq 0, \forall k \in [K], \forall r \in [R_k], \\
&0 \leq w_{ikr} \leq \hat{E}_{1ikr}^\ell, \forall r \in [R_k], \forall k \in [K], \\
&0 \leq u_{ikr} - w_{ikr} \leq (1 - \hat{E}_{1ikr}^\ell), \forall r \in [R_k], \forall k \in [K], \\
&\sum_{\tau \in [R_k]} 2^{r-1} \delta_k c_k w_{ikr} \geq \tilde{L}_{ijk}, \forall k \in [K], \\
&\sum_{k \in [K]} \tilde{L}_{ijk}^\ell = L_{ij}^\ell, 0 \leq \hat{E}_{1ik}^\ell \leq E_{1ij}^\ell \leq 1, \tilde{L}_{ijk}^\ell \geq 0, \forall k \in [K] \end{align*} \right\}.
\]

(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:

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

(22)

Under-capacity Inequalities: Let us define \( M = \min_{k \in [K]} [\delta_k c_k] \) 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:

\[
\ell_{ij}^\ell = \left[ (1 - \hat{a}_{ij}^\ell) \ell_{ij-1}^\ell \right] + \hat{\lambda}_{ij}^\ell, \forall i \in [I], \forall j \in [J], \forall \ell \in [N],
\]

(23)

where we let \( \ell_0 = 0 \) for all \( i \in [I] \) and \( \ell \in [N]. \) Now let us define \( J_{is}^\ell = \min\{J, \min_{j \in [J]} \{\ell_{i0}^\ell > M\} \} \) 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 \( j \in [J_{is}^\ell - 1], \) it is under-capacity, i.e., we must have

\[
y_{ij}^\ell = 1, \forall i \in [I], \forall j \in [J_{is}^\ell - 1], \forall \ell \in [N].
\]

(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:

\[
1 - y_{ij}^\ell \leq E_{1ij}^\ell, \forall i \in [I], \forall j \in [J_{is}^\ell], \forall \ell \in [N].
\]

(25)
Tightening Big-M Coefficients: Note that in (10), we must have $L_{ij}^t \geq \tilde{\lambda}_{ij}^t$ if $\tilde{y}_{ij}^t = 1$ and $L_{ij}^t \geq \tilde{L}_{i,j-1}^t$. Hence, in the first part of (10), we can reduce $M$ by $M - \tilde{\lambda}_{ij}^t$, and in the second part of (10), $\tilde{\lambda}_{ij}^t - M(1 - y_{ij}^t)$ can be tightened by $\tilde{\lambda}_{ij}^t y_{ij}^t$. Therefore, we arrive at the following stronger inequalities

$$L_{ij}^t \geq \sum_{k \in [K]} n_{ik}[\delta_k c_k] - (M - \tilde{\lambda}_{ij}^t) y_{ij}^t, L_{ij}^t \geq \tilde{L}_{i,j-1}^t + \tilde{\lambda}_{ij}^t y_{ij}^t, \forall i \in [I], \forall j \in [J], \forall \ell \in [N].$$ (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

$$E_{ij}^t \leq E_{i}^t, \forall \ell \in [N].$$ (27)

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

$$\tilde{E}_{i}^t \geq \frac{\tilde{L}_{i,jk}^t}{\tilde{M}_k[\delta_k c_k]}, \forall i \in [I], \forall j \in [J], \forall \ell \in [N].$$ (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 $u$ and the linearization of constraints (2b) using auxiliary variables $y$. 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 $\ell$ 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

$$\min_{n,x} \left\{ \frac{1}{N} \sum_{\ell \in [N]} \mathcal{Q}(n, x, \xi^\ell) : (1b) - (1f) \right\}.$$ (29a)

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

$$\mathcal{Q}(n, x, \xi^\ell) = \min_{E^\ell \in \mathbb{R}_+} \left\{ E^\ell : \left\| 2 \sqrt{\max_{j \in [J]} \ell_{ij}^\ell} \sum_{k \in [K]} n_{ik}[\delta_k c_k] - E^\ell \right\|_2 \leq \sum_{k \in [K]} n_{ik}[\delta_k c_k] + E^\ell, \forall i \in [I] \right\}.$$ (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 \((\bar{n}, \bar{x})\) into DrFRAM (6), i.e.,

\[
v^N = \frac{1}{N} \sum_{\ell \in [N]} Q(\bar{n}, \bar{x}, \xi^\ell).
\]

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

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

\[
v^* \leq v^N \leq v^* + \frac{\omega}{N} \| (\bar{E} - e) + \| + \frac{M/M - 1}{N} \| E_2^* \|_0,
\]

where \(M := \min_{k \in \{1, \ldots, M\}} |\delta_k c_k|\) and \(\bar{M} := \max_{i \in \{1, \ldots, I\}, j \in \{1, \ldots, J\}, \ell \in \{1, \ldots, \ell\} |\ell_i^j|\).

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

Next, we note that if \(\bar{E}^\ell > 1\), then we have \(1 \leq Q(\bar{n}, \bar{x}, \xi^\ell) \leq 1 + \omega\). Therefore, we must have

\[
v^N - \frac{\omega}{N} \| (\bar{E} - e) + \| \leq \frac{1}{N} \sum_{\ell \in [N]} E_\ell.
\]

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

\[
\frac{1}{N} \sum_{\ell \in [N]} E_\ell \leq \frac{1}{N} \sum_{\ell \in [N]} \max_{i \in \{1, \ldots, I\}, j \in \{1, \ldots, J\}, \ell \in \{1, \ldots, \ell\} |\ell_i^j| \leq v^* + \frac{\bar{M}/M - 1}{N} \| E_2^* \|_0.
\]

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

Proposition 8 shows that \((\bar{n}, \bar{x})\) is optimal to DrFRAM (6) if no passenger is being abandoned (i.e., \(E_2^* = 0\)) 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 \((\tilde{n}, \tilde{x})\) denotes an optimal first-stage decision of the No-one-left model (29). Let us denote two subsets of scenarios, \(N_- = \{\ell \in [N] : Q(\tilde{n}, \tilde{x}, \xi^\ell) \leq 1\}\) and \(N_+ = \{\ell \in [N] : Q(\tilde{n}, \tilde{x}, \xi^\ell) > 1\}\). Then the enhanced model is defined as follows:

\[
(\text{EnModel}) \quad \min_{n, x} \left\{ \frac{1}{N} \left[ \sum_{\ell \in N_-} Q(n, x, \xi^\ell) + \sum_{\ell \in N_+} Q^E(n, x, \xi^\ell) \right] : (1b) - (1f) \right\},
\]

where \(Q(n, x, \xi^\ell)\) is defined in (29) and

\[
Q^E(n, x, \xi^\ell) = \min_{L^\ell, L^{\ell'}, g, E_2} \left\{ 1 + E_2^\ell : (9), (10), (11b), E_2^\ell \in [0, 1] \right\}.
\]

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

1. Initialization: Solve the No-one-left model (29) with an optimal first-stage solution \((\bar{n}, \bar{x})\);
   set \(t = 0, N_- = [N]\) and initialize \(t_{\text{max}}\);

2. while Set \(N_-\) is changing and \(t < t_{\text{max}}\) do
   3. Define \(N_- = \{\ell \in [N] : Q(\bar{n}, \bar{x}, \hat{\xi}^\ell) \leq 1\}\) and \(N_+ = \{\ell \in [N] : Q(\bar{n}, \bar{x}, \hat{\xi}^\ell) > 1\}\);
   4. Solve the EnModel (32);
   5. \(t \leftarrow t + 1\).

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 processor 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 \(I, \eta, \delta, \omega, \theta\)). 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 $\theta$ increases, the objective value increases, and solution time does not change too much, since any increase in $\theta$ may result in larger passenger arrival rates and smaller passenger alighting rates. As the pandemic factor $\delta$ or the total number of buses increases and other parameters stay the same, both objective value and solution time decrease, since larger $\delta$ or larger $\eta$ implies larger capacities for some routes and thus more passengers to get onto the bus. Weight $\omega$ 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 $\omega$ 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**

<table>
<thead>
<tr>
<th>Case</th>
<th>$I$</th>
<th>$\eta$</th>
<th>$\delta$</th>
<th>$\omega$</th>
<th>$\theta$</th>
<th>OPT</th>
<th>MILP.B</th>
<th>MILP.VI</th>
<th>MILP.CONV</th>
<th>MILP.VI.CONV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.25</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td>49</td>
<td>40</td>
<td>26</td>
<td>25</td>
</tr>
<tr>
<td>1 - 1</td>
<td>0</td>
<td>0.88</td>
<td>64</td>
<td>73</td>
<td>24</td>
<td>28</td>
<td>2883</td>
<td>30</td>
<td>2684</td>
<td>2524</td>
</tr>
<tr>
<td>1 - 2</td>
<td>0.25</td>
<td>0.92</td>
<td>56</td>
<td>128</td>
<td>67</td>
<td>70</td>
<td>5123</td>
<td>14</td>
<td>10032</td>
<td>10385</td>
</tr>
<tr>
<td>1 - 3</td>
<td>0.25</td>
<td>0.25</td>
<td>36</td>
<td>93</td>
<td>70</td>
<td>72</td>
<td>3781</td>
<td>40</td>
<td>7375</td>
<td>9843</td>
</tr>
<tr>
<td>1 - 4</td>
<td>0.25</td>
<td>0.25</td>
<td>73</td>
<td>182</td>
<td>30</td>
<td>32</td>
<td>7065</td>
<td>75</td>
<td>7385</td>
<td>4823</td>
</tr>
<tr>
<td>1 - 5</td>
<td>1</td>
<td>0.94</td>
<td>1.5</td>
<td>9.72</td>
<td>45</td>
<td>58</td>
<td>7580</td>
<td>28</td>
<td>6732</td>
<td>4912</td>
</tr>
<tr>
<td>Average</td>
<td>0</td>
<td>0.88</td>
<td>74</td>
<td>185</td>
<td>59</td>
<td>81</td>
<td>4154</td>
<td>60</td>
<td>5164</td>
<td>3012</td>
</tr>
<tr>
<td>1 - 6</td>
<td>0</td>
<td>0.88</td>
<td>64</td>
<td>73</td>
<td>24</td>
<td>28</td>
<td>2883</td>
<td>30</td>
<td>2659</td>
<td>1911</td>
</tr>
<tr>
<td>1 - 7</td>
<td>0.25</td>
<td>0.92</td>
<td>56</td>
<td>128</td>
<td>67</td>
<td>70</td>
<td>5123</td>
<td>14</td>
<td>10032</td>
<td>10385</td>
</tr>
<tr>
<td>1 - 8</td>
<td>0.25</td>
<td>0.25</td>
<td>36</td>
<td>93</td>
<td>70</td>
<td>72</td>
<td>3781</td>
<td>40</td>
<td>7375</td>
<td>9843</td>
</tr>
<tr>
<td>1 - 9</td>
<td>0.25</td>
<td>0.25</td>
<td>73</td>
<td>182</td>
<td>30</td>
<td>32</td>
<td>7065</td>
<td>75</td>
<td>7385</td>
<td>4823</td>
</tr>
<tr>
<td>Average</td>
<td>0</td>
<td>0.88</td>
<td>74</td>
<td>185</td>
<td>59</td>
<td>81</td>
<td>4154</td>
<td>60</td>
<td>5164</td>
<td>3012</td>
</tr>
<tr>
<td>1 - 11</td>
<td>0</td>
<td>0.572</td>
<td>20</td>
<td>100</td>
<td>10</td>
<td>53</td>
<td>1955</td>
<td>18</td>
<td>9968</td>
<td>2899</td>
</tr>
<tr>
<td>1 - 12</td>
<td>0</td>
<td>0.596</td>
<td>24</td>
<td>128</td>
<td>24</td>
<td>12</td>
<td>3101</td>
<td>21</td>
<td>2496</td>
<td>1402</td>
</tr>
<tr>
<td>1 - 13</td>
<td>0</td>
<td>0.572</td>
<td>24</td>
<td>100</td>
<td>10</td>
<td>53</td>
<td>1955</td>
<td>18</td>
<td>9968</td>
<td>2899</td>
</tr>
<tr>
<td>1 - 14</td>
<td>0</td>
<td>0.596</td>
<td>24</td>
<td>128</td>
<td>24</td>
<td>12</td>
<td>3101</td>
<td>21</td>
<td>2496</td>
<td>1402</td>
</tr>
<tr>
<td>1 - 15</td>
<td>0</td>
<td>0.572</td>
<td>24</td>
<td>100</td>
<td>10</td>
<td>53</td>
<td>1955</td>
<td>18</td>
<td>9968</td>
<td>2899</td>
</tr>
<tr>
<td>Average</td>
<td>0</td>
<td>0.572</td>
<td>20</td>
<td>100</td>
<td>10</td>
<td>53</td>
<td>1955</td>
<td>18</td>
<td>9968</td>
<td>2899</td>
</tr>
<tr>
<td>1 - 16</td>
<td>0</td>
<td>0.288</td>
<td>3</td>
<td>185</td>
<td>53</td>
<td>74</td>
<td>1562</td>
<td>4</td>
<td>3518</td>
<td>1125</td>
</tr>
<tr>
<td>1 - 17</td>
<td>0</td>
<td>0.25</td>
<td>6</td>
<td>2957</td>
<td>36</td>
<td>8</td>
<td>2957</td>
<td>36</td>
<td>8268</td>
<td>1125</td>
</tr>
<tr>
<td>1 - 18</td>
<td>0</td>
<td>0.25</td>
<td>6</td>
<td>2957</td>
<td>36</td>
<td>8</td>
<td>2957</td>
<td>36</td>
<td>8268</td>
<td>1125</td>
</tr>
<tr>
<td>1 - 19</td>
<td>0</td>
<td>0.25</td>
<td>6</td>
<td>2957</td>
<td>36</td>
<td>8</td>
<td>2957</td>
<td>36</td>
<td>8268</td>
<td>1125</td>
</tr>
<tr>
<td>1 - 20</td>
<td>0</td>
<td>0.25</td>
<td>6</td>
<td>2957</td>
<td>36</td>
<td>8</td>
<td>2957</td>
<td>36</td>
<td>8268</td>
<td>1125</td>
</tr>
<tr>
<td>Average</td>
<td>0</td>
<td>0.288</td>
<td>3</td>
<td>185</td>
<td>53</td>
<td>74</td>
<td>1562</td>
<td>4</td>
<td>3518</td>
<td>1125</td>
</tr>
<tr>
<td>2 - 26</td>
<td>0</td>
<td>0.389</td>
<td>5</td>
<td>7389</td>
<td>26</td>
<td>36</td>
<td>10541</td>
<td>16</td>
<td>4261</td>
<td>11674</td>
</tr>
<tr>
<td>2 - 27</td>
<td>0</td>
<td>0.404</td>
<td>111</td>
<td>12592</td>
<td>31</td>
<td>15186</td>
<td>440</td>
<td>8289</td>
<td>8241</td>
<td>11568</td>
</tr>
<tr>
<td>2 - 28</td>
<td>0</td>
<td>0.404</td>
<td>111</td>
<td>12592</td>
<td>31</td>
<td>15186</td>
<td>440</td>
<td>8289</td>
<td>8241</td>
<td>11568</td>
</tr>
<tr>
<td>2 - 29</td>
<td>0</td>
<td>0.404</td>
<td>111</td>
<td>12592</td>
<td>31</td>
<td>15186</td>
<td>440</td>
<td>8289</td>
<td>8241</td>
<td>11568</td>
</tr>
<tr>
<td>2 - 30</td>
<td>0</td>
<td>0.404</td>
<td>111</td>
<td>12592</td>
<td>31</td>
<td>15186</td>
<td>440</td>
<td>8289</td>
<td>8241</td>
<td>11568</td>
</tr>
<tr>
<td>Average</td>
<td>0</td>
<td>0.389</td>
<td>5</td>
<td>7389</td>
<td>26</td>
<td>36</td>
<td>10541</td>
<td>16</td>
<td>4261</td>
<td>11674</td>
</tr>
<tr>
<td>2 - 31</td>
<td>0</td>
<td>0.389</td>
<td>5</td>
<td>7389</td>
<td>26</td>
<td>36</td>
<td>10541</td>
<td>16</td>
<td>4261</td>
<td>11674</td>
</tr>
<tr>
<td>2 - 32</td>
<td>0</td>
<td>0.404</td>
<td>111</td>
<td>12592</td>
<td>31</td>
<td>15186</td>
<td>440</td>
<td>8289</td>
<td>8241</td>
<td>11568</td>
</tr>
<tr>
<td>2 - 33</td>
<td>0</td>
<td>0.404</td>
<td>111</td>
<td>12592</td>
<td>31</td>
<td>15186</td>
<td>440</td>
<td>8289</td>
<td>8241</td>
<td>11568</td>
</tr>
<tr>
<td>2 - 34</td>
<td>0</td>
<td>0.404</td>
<td>111</td>
<td>12592</td>
<td>31</td>
<td>15186</td>
<td>440</td>
<td>8289</td>
<td>8241</td>
<td>11568</td>
</tr>
<tr>
<td>2 - 35</td>
<td>0</td>
<td>0.404</td>
<td>111</td>
<td>12592</td>
<td>31</td>
<td>15186</td>
<td>440</td>
<td>8289</td>
<td>8241</td>
<td>11568</td>
</tr>
</tbody>
</table>
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>
<thead>
<tr>
<th>Case</th>
<th>C.MILP.B</th>
<th>C.MILP.VI</th>
<th>C.MILP.CONV</th>
<th>C.MILP.VI.CONV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obj Val</td>
<td>Time</td>
<td>Gap (%)</td>
<td>Obj Val</td>
<td>Time</td>
</tr>
<tr>
<td>1 - 1</td>
<td>0.000</td>
<td>0.02</td>
<td>100.0</td>
<td>0.000</td>
</tr>
<tr>
<td>1 - 2</td>
<td>0.000</td>
<td>0.02</td>
<td>100.0</td>
<td>0.091</td>
</tr>
<tr>
<td>1 - 3</td>
<td>0.000</td>
<td>0.02</td>
<td>100.0</td>
<td>0.093</td>
</tr>
<tr>
<td>1 - 4</td>
<td>0.000</td>
<td>0.02</td>
<td>100.0</td>
<td>0.094</td>
</tr>
<tr>
<td>1 - 5</td>
<td>0.000</td>
<td>0.02</td>
<td>100.0</td>
<td>0.098</td>
</tr>
<tr>
<td>Average</td>
<td>0.000</td>
<td>0.02</td>
<td>100.0</td>
<td>0.089</td>
</tr>
</tbody>
</table>

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.
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 $\theta$ increases, the solution time does not change too much for all the scenario decomposition methods. As the pandemic factor $\delta$ or the number of buses $\eta$ increases, the solution time decreases. As the weight $\omega$ 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)

<table>
<thead>
<tr>
<th>Case</th>
<th>MILP.B</th>
<th>MILP.VI</th>
<th>MILP.VI.COM</th>
<th>Approximate Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time</td>
<td>Gap (%)</td>
<td>Time</td>
<td>Algorithm 3</td>
</tr>
<tr>
<td>1 - 1</td>
<td>22</td>
<td>0.0</td>
<td>8.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1 - 2</td>
<td>14</td>
<td>0.0</td>
<td>6.0</td>
<td>28.0</td>
</tr>
<tr>
<td>1 - 3</td>
<td>27</td>
<td>0.0</td>
<td>9.0</td>
<td>36.0</td>
</tr>
<tr>
<td>1 - 4</td>
<td>40</td>
<td>0.0</td>
<td>8.0</td>
<td>128.0</td>
</tr>
<tr>
<td>1 - 5</td>
<td>72</td>
<td>0.0</td>
<td>8.0</td>
<td>83.0</td>
</tr>
<tr>
<td>Average</td>
<td>33</td>
<td>0.0</td>
<td>8.0</td>
<td>62.0</td>
</tr>
<tr>
<td></td>
<td>1 - 6</td>
<td>35</td>
<td>0.0</td>
<td>39.0</td>
</tr>
<tr>
<td></td>
<td>1 - 7</td>
<td>19</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>1 - 8</td>
<td>50</td>
<td>0.0</td>
<td>34.0</td>
</tr>
<tr>
<td></td>
<td>1 - 9</td>
<td>23</td>
<td>0.0</td>
<td>49.0</td>
</tr>
<tr>
<td></td>
<td>1 - 10</td>
<td>22</td>
<td>0.0</td>
<td>17.0</td>
</tr>
<tr>
<td>Average</td>
<td>29</td>
<td>0.0</td>
<td>10.0</td>
<td>45.0</td>
</tr>
<tr>
<td></td>
<td>1 - 11</td>
<td>21</td>
<td>0.0</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td>1 - 12</td>
<td>17</td>
<td>0.0</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>1 - 13</td>
<td>15</td>
<td>0.0</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>1 - 14</td>
<td>19</td>
<td>0.0</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td>1 - 15</td>
<td>49</td>
<td>0.0</td>
<td>13.0</td>
</tr>
<tr>
<td>Average</td>
<td>24</td>
<td>0.0</td>
<td>13.0</td>
<td>47.0</td>
</tr>
<tr>
<td></td>
<td>1 - 16</td>
<td>39</td>
<td>0.0</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>1 - 17</td>
<td>25</td>
<td>0.0</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>1 - 18</td>
<td>27</td>
<td>0.0</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>1 - 19</td>
<td>33</td>
<td>0.0</td>
<td>9.0</td>
</tr>
<tr>
<td></td>
<td>1 - 20</td>
<td>33</td>
<td>0.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Average</td>
<td>31</td>
<td>0.0</td>
<td>10.0</td>
<td>58.0</td>
</tr>
<tr>
<td></td>
<td>1 - 21</td>
<td>20</td>
<td>0.0</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>1 - 22</td>
<td>14</td>
<td>0.0</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>1 - 23</td>
<td>12</td>
<td>0.0</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>1 - 24</td>
<td>15</td>
<td>0.0</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>1 - 25</td>
<td>13</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td>Average</td>
<td>25</td>
<td>0.0</td>
<td>8.0</td>
<td>45.0</td>
</tr>
<tr>
<td></td>
<td>2 - 26</td>
<td>38</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>2 - 27</td>
<td>44</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>2 - 28</td>
<td>41</td>
<td>0.0</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>2 - 29</td>
<td>35</td>
<td>0.0</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>2 - 30</td>
<td>25</td>
<td>0.0</td>
<td>12.0</td>
</tr>
<tr>
<td>Average</td>
<td>27</td>
<td>0.0</td>
<td>12.0</td>
<td>62.0</td>
</tr>
<tr>
<td></td>
<td>2 - 31</td>
<td>35</td>
<td>0.0</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>2 - 32</td>
<td>34</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>2 - 33</td>
<td>30</td>
<td>0.0</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>2 - 34</td>
<td>27</td>
<td>0.0</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>2 - 35</td>
<td>21</td>
<td>0.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Average</td>
<td>29</td>
<td>0.0</td>
<td>8.0</td>
<td>55.0</td>
</tr>
</tbody>
</table>

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 $\delta$ is 0.25. We tested on 20 cases based on these instances using different parameter configurations.
Table 4: Results of Instances with a Larger Objective Value

<table>
<thead>
<tr>
<th>Case</th>
<th>N</th>
<th>I</th>
<th>η</th>
<th>ω</th>
<th>MILP.B</th>
<th>MILP.VI</th>
<th>Algorithm 1.VI</th>
<th>Algorithm 3</th>
<th>Algorithm 4 (max itr 5)</th>
<th>Algorithm 4 (max itr 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 - 36</td>
<td>0.25</td>
<td>0</td>
<td>1.006</td>
<td>56</td>
<td>25</td>
<td>0.0</td>
<td>11</td>
<td>0.0</td>
<td>12</td>
<td>0.0</td>
</tr>
<tr>
<td>3 - 37</td>
<td>0.25</td>
<td>1.053</td>
<td>33</td>
<td>46</td>
<td>0.0</td>
<td>15</td>
<td>0.0</td>
<td>16</td>
<td>0.0</td>
<td>1.063</td>
</tr>
<tr>
<td>3 - 38</td>
<td>0.5</td>
<td>1.053</td>
<td>19</td>
<td>54</td>
<td>0.0</td>
<td>24</td>
<td>0.0</td>
<td>19</td>
<td>0.0</td>
<td>1.063</td>
</tr>
<tr>
<td>3 - 39</td>
<td>1</td>
<td>1.053</td>
<td>35</td>
<td>31</td>
<td>0.0</td>
<td>39</td>
<td>0.0</td>
<td>23</td>
<td>0.0</td>
<td>1.063</td>
</tr>
<tr>
<td>3 - 40</td>
<td>1.5</td>
<td>1.076</td>
<td>33</td>
<td>12</td>
<td>0.0</td>
<td>23</td>
<td>0.0</td>
<td>41</td>
<td>0.0</td>
<td>1.076</td>
</tr>
<tr>
<td>Average</td>
<td>35</td>
<td>34</td>
<td>0.0</td>
<td>22</td>
<td>0.0</td>
<td>22</td>
<td>0.0</td>
<td>4</td>
<td>1.2</td>
<td>11</td>
</tr>
<tr>
<td>4 - 44</td>
<td>0.5</td>
<td>0</td>
<td>1.032</td>
<td>24</td>
<td>41</td>
<td>0.0</td>
<td>15</td>
<td>0.0</td>
<td>22</td>
<td>0.0</td>
</tr>
<tr>
<td>4 - 45</td>
<td>0.25</td>
<td>1.105</td>
<td>48</td>
<td>24</td>
<td>0.0</td>
<td>33</td>
<td>0.0</td>
<td>24</td>
<td>0.0</td>
<td>1.135</td>
</tr>
<tr>
<td>4 - 43</td>
<td>0.5</td>
<td>1.105</td>
<td>40</td>
<td>19</td>
<td>0.0</td>
<td>26</td>
<td>0.0</td>
<td>17</td>
<td>0.0</td>
<td>1.135</td>
</tr>
<tr>
<td>4 - 44</td>
<td>1</td>
<td>1.105</td>
<td>23</td>
<td>25</td>
<td>0.0</td>
<td>21</td>
<td>0.0</td>
<td>20</td>
<td>0.0</td>
<td>1.135</td>
</tr>
<tr>
<td>4 - 45</td>
<td>1.5</td>
<td>1.152</td>
<td>35</td>
<td>11</td>
<td>0.0</td>
<td>24</td>
<td>0.0</td>
<td>23</td>
<td>0.0</td>
<td>1.152</td>
</tr>
<tr>
<td>Average</td>
<td>34</td>
<td>24</td>
<td>0.0</td>
<td>24</td>
<td>0.0</td>
<td>21</td>
<td>0.0</td>
<td>4</td>
<td>2.1</td>
<td>18</td>
</tr>
<tr>
<td>5 - 46</td>
<td>1</td>
<td>0.105</td>
<td>33</td>
<td>15</td>
<td>0.0</td>
<td>15</td>
<td>0.0</td>
<td>11</td>
<td>0.0</td>
<td>1.093</td>
</tr>
<tr>
<td>5 - 47</td>
<td>0.25</td>
<td>1.210</td>
<td>29</td>
<td>13</td>
<td>0.0</td>
<td>19</td>
<td>0.0</td>
<td>14</td>
<td>0.0</td>
<td>1.231</td>
</tr>
<tr>
<td>5 - 48</td>
<td>0.5</td>
<td>1.210</td>
<td>36</td>
<td>34</td>
<td>0.0</td>
<td>17</td>
<td>0.0</td>
<td>18</td>
<td>0.0</td>
<td>1.231</td>
</tr>
<tr>
<td>5 - 49</td>
<td>1</td>
<td>1.210</td>
<td>25</td>
<td>14</td>
<td>0.0</td>
<td>17</td>
<td>0.0</td>
<td>18</td>
<td>0.0</td>
<td>1.231</td>
</tr>
<tr>
<td>5 - 50</td>
<td>1.5</td>
<td>1.303</td>
<td>36</td>
<td>36</td>
<td>0.0</td>
<td>21</td>
<td>0.0</td>
<td>15</td>
<td>0.0</td>
<td>1.333</td>
</tr>
<tr>
<td>Average</td>
<td>34</td>
<td>22</td>
<td>0.0</td>
<td>18</td>
<td>0.0</td>
<td>14</td>
<td>0.0</td>
<td>7</td>
<td>1.6</td>
<td>17</td>
</tr>
</tbody>
</table>

The optimality gap is quite large within the time limit. However, albeit promising, MILP.VI may not be ideal for solving extremely large instances since 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
Table 6  Results of MILP.CONV and MILP.VI.CONV for Solving Larger Instances

<table>
<thead>
<tr>
<th>Case</th>
<th>MILP.CONV</th>
<th>MILP.VI.CONV</th>
<th>MILP.VI.CONV</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 - 51</td>
<td>0.755</td>
<td>3600</td>
<td>49.5</td>
</tr>
<tr>
<td>6 - 52</td>
<td>0.773</td>
<td>3600</td>
<td>48.2</td>
</tr>
<tr>
<td>6 - 53</td>
<td>0.787</td>
<td>3600</td>
<td>42.9</td>
</tr>
<tr>
<td>6 - 54</td>
<td>0.788</td>
<td>3600</td>
<td>55.4</td>
</tr>
<tr>
<td>6 - 55</td>
<td>0.861</td>
<td>3600</td>
<td>54.1</td>
</tr>
</tbody>
</table>

Average: 3600 44.8 3141 3600 51.1 15286

6 - 56 | 0.767 | 3600 | 45.4 | 0.419 | 20850 |
| 6 - 57 | 0.773 | 3600 | 48.2 | 0.400 | 1352 |
| 6 - 58 | 0.847 | 3600 | 51.3 | 0.413 | 21090 |
| 6 - 59 | 0.864 | 3600 | 54.1 | 0.395 | 2274 |

Average: 3600 49.5 11302

6 - 60 | 0.767 | 3600 | 45.4 | 0.419 | 20850 |
| 6 - 61 | 0.773 | 3600 | 48.2 | 0.400 | 1352 |
| 6 - 62 | 0.847 | 3600 | 51.3 | 0.413 | 21090 |
| 6 - 63 | 0.864 | 3600 | 54.1 | 0.395 | 2274 |

Average: 3600 49.5 11302

6 - 64 | 0.787 | 3600 | 42.9 | 0.449 | 21090 |
| 6 - 65 | 0.788 | 3600 | 55.4 | 0.352 | 5154 |
| 6 - 66 | 0.861 | 3600 | 54.1 | 0.395 | 2274 |
| 6 - 67 | 0.864 | 3600 | 54.1 | 0.370 | 15286 |

Average: 3600 51.1 15286

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

<table>
<thead>
<tr>
<th>Case</th>
<th>C.MILP.B</th>
<th>C.MILP.VI</th>
<th>C.MILP.CONV</th>
<th>C.MILP.VI.CONV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obj.Val</td>
<td>Time (s)</td>
<td>Opt.Gap (%)</td>
<td>LB Nodes</td>
</tr>
<tr>
<td>6 - 51</td>
<td>0.000</td>
<td>0.16</td>
<td>0.044</td>
<td>0.85</td>
</tr>
<tr>
<td>6 - 52</td>
<td>0.000</td>
<td>0.15</td>
<td>0.045</td>
<td>0.96</td>
</tr>
<tr>
<td>6 - 53</td>
<td>0.000</td>
<td>0.16</td>
<td>0.046</td>
<td>0.58</td>
</tr>
<tr>
<td>6 - 54</td>
<td>0.000</td>
<td>0.34</td>
<td>0.046</td>
<td>1.09</td>
</tr>
<tr>
<td>6 - 55</td>
<td>0.000</td>
<td>0.19</td>
<td>0.048</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Average: 0.20 0.85 0.41 2.95

6 - 56 | 0.000 | 0.16 | 0.044 | 0.98 | 0.030 | 0.46 | 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: 0.22 2.18 1.40 6.76

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 | 0.003 | 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 \(P^\infty\) denote the sample probability distribution, and \(P^*\) be the true distribution of random parameters \(\tilde{\xi}\). Motivated by Esfahani and Kuhn (2018), we define the out-of-sample probability as

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

where \((n^D, x^D)\) denotes the optimal solution of the DrFRAM model and \(Q(\cdot, \cdot, \cdot)\) 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 \(2\alpha\); 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

<table>
<thead>
<tr>
<th>Case</th>
<th>MILP.B</th>
<th>MILP.CONV</th>
<th>MILP.VI.CONV</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 - 51</td>
<td>0.743</td>
<td>0.768</td>
<td>3600</td>
</tr>
<tr>
<td>6 - 52</td>
<td>0.763</td>
<td>0.734</td>
<td>3600</td>
</tr>
<tr>
<td>6 - 53</td>
<td>0.775</td>
<td>0.745</td>
<td>3600</td>
</tr>
<tr>
<td>6 - 54</td>
<td>0.778</td>
<td>0.748</td>
<td>3600</td>
</tr>
<tr>
<td>6 - 55</td>
<td>0.817</td>
<td>0.785</td>
<td>3600</td>
</tr>
</tbody>
</table>

Average | 3600 | 4.7 | 3600 | 3.3 |

7 - 61 | 3600 | / | / | / | 3600 | / | / | / | 3600 | / | / | / | 3600 |
7 - 62 | / | / | 3600 | / | / | / | / | / | / | / | / | / | / | / |
7 - 63 | / | / | 3600 | / | / | / | / | / | / | / | / | / | / | / |
7 - 64 | / | / | 3600 | / | / | / | / | / | / | / | / | / | / | / |
7 - 65 | / | / | 3600 | / | / | / | / | / | / | / | / | / | / | / |
Average | 3600 | / | / | / | 3600 | / | / | / | 3600 | / | / | / | 3600 |

8 - 66 | / | / | 3600 | / | / | / | / | / | / | / | / | / | / | / |
8 - 67 | / | / | 3600 | / | / | / | / | / | / | / | / | / | / | / |
8 - 68 | / | / | 3600 | / | / | / | / | / | / | / | / | / | / | / |
8 - 69 | / | / | 3600 | / | / | / | / | / | / | / | / | / | / | / |
8 - 70 | / | / | 3600 | / | / | / | / | / | / | / | / | / | / | / |
Average | 3600 | / | / | / | 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

<table>
<thead>
<tr>
<th>Case</th>
<th>Algorithm 1.VI</th>
<th>Algorithm 2.VI</th>
<th>Best.Obj</th>
<th>Algorithm 3</th>
<th>Algorithm 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 - 51</td>
<td>0.743</td>
<td>0.719</td>
<td>3600</td>
<td>4.3</td>
<td>0.743</td>
</tr>
<tr>
<td>6 - 52</td>
<td>0.763</td>
<td>0.748</td>
<td>3600</td>
<td>2.0</td>
<td>0.763</td>
</tr>
<tr>
<td>6 - 53</td>
<td>0.775</td>
<td>0.757</td>
<td>3600</td>
<td>2.2</td>
<td>0.775</td>
</tr>
<tr>
<td>6 - 54</td>
<td>0.778</td>
<td>0.762</td>
<td>3600</td>
<td>2.0</td>
<td>0.778</td>
</tr>
<tr>
<td>6 - 55</td>
<td>0.817</td>
<td>0.797</td>
<td>3600</td>
<td>2.5</td>
<td>0.817</td>
</tr>
</tbody>
</table>

Average | 3600 | 2.3 | 3600 | 4.3 | 3600 | 2.1 | 3600 | 2.2 | 133 | 0.92 |

7 - 61 | / | / | 3600 | / | / | / | / | / | 1.249 | / | 3600 | 1.124 | 3600 |
7 - 62 | / | / | 3600 | / | / | / | / | / | 1.300 | / | 3600 | 1.139 | 3600 |
7 - 63 | / | / | 3600 | / | / | / | / | / | 1.272 | / | 3600 | 1.122 | 3600 |
7 - 64 | / | / | 3600 | / | / | / | / | / | 1.316 | / | 3600 | 1.148 | 3600 |
7 - 65 | / | / | 3600 | / | / | / | / | / | 1.299 | / | 3600 | 1.163 | 3600 |

Average | 3600 | 2.1 | 3600 | 2.2 | 3600 | 2.1 | 3600 | 2.1 | 133 | 0.92 |

8 - 66 | 0.768   | 0.695 | 3600 | 9.4 | / | / | / | / | 0.798 | / | 3600 | 0.764 | 776 |
8 - 67 | 0.789   | 0.718 | 3600 | 9.0 | / | / | / | / | 0.923 | / | 3600 | 0.780 | 270 |
8 - 68 | 0.800   | 0.723 | 3600 | 9.7 | / | / | / | / | 1.349 | / | 3600 | 0.789 | 390 |
8 - 69 | 0.819   | 0.735 | 3600 | 10.3 | / | / | / | / | 1.294 | / | 3600 | 0.800 | 349 |
8 - 70 | 0.849   | 0.769 | 3600 | 9.4 | / | / | / | / | 0.867 | / | 3600 | 0.839 | 888 |

Average | 3600 | 9.6 | 3600 | 9.5 | 3600 | 9.5 | 3600 | 9.5 | 265 | 2.4 |

θ, we first solved the DrFRAM model using generated data and then generated new samples with the same sample size and obtained (1 − α) 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 c₁ = 60, c₂ = 80, c₃ = 120, the
number of buses is $\eta = (6, 9, 3)$, the pandemic factor $\delta$ is 0.25, the weight $\omega$ 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 $\theta$ 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 $\theta$ 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.](image-url)
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 $\delta$ 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 $\theta$ for a given $\delta$ 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 $\theta$ 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 $\theta$ 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 $\theta$ 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

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\theta$</th>
<th>95% C.I. of DrFRAM</th>
<th>95% C.I. of Mean-Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0</td>
<td>[0.122, 0.135]</td>
<td>[0.219, 0.254]</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>[0.059, 0.080]</td>
<td>[0.555, 0.590]</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>[0.565, 0.586]</td>
<td>[0.559, 0.594]</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>[0.571, 0.593]</td>
<td>[0.563, 0.598]</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>[0.578, 0.600]</td>
<td>[0.569, 0.603]</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>[0.588, 0.612]</td>
<td>[0.578, 0.616]</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>[1.051, 1.131]</td>
<td>[0.845, 1.001]</td>
</tr>
<tr>
<td></td>
<td>1.4</td>
<td>[1.130, 1.217]</td>
<td>[1.006, 1.129]</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>[1.134, 1.224]</td>
<td>[1.001, 1.163]</td>
</tr>
<tr>
<td></td>
<td>1.8</td>
<td>[1.073, 1.154]</td>
<td>[0.865, 0.984]</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>[1.016, 1.075]</td>
<td>[0.784, 0.873]</td>
</tr>
<tr>
<td></td>
<td>2.2</td>
<td>[1.412, 1.452]</td>
<td>[1.081, 1.090]</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>[1.418, 1.451]</td>
<td>[1.070, 1.082]</td>
</tr>
<tr>
<td></td>
<td>2.6</td>
<td>[1.377, 1.418]</td>
<td>[1.099, 1.114]</td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>[1.427, 1.454]</td>
<td>[1.083, 1.089]</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>[1.430, 1.462]</td>
<td>[1.084, 1.098]</td>
</tr>
</tbody>
</table>

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 $\delta$ and weight $\omega$ by varying $\delta$ from 0.25 to 1 and $\omega$ from 0.25 to 2. While varying $\delta$, we let $\omega = 0.5$; while varying $\omega$, we let $\delta = 0.25$. We tuned $\theta$ 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 $\theta$ to derive 95% confidence intervals. The results are illustrated in Figure 4. It is seen that the objective value
gets smaller when the pandemic factor $\delta$ 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 $\omega$ increases since a higher weight $\omega$ leads to a larger objective value.

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](image)

**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-$\infty$ 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
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


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:

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

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

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

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

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

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). \( \square \)
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 \( \mu = (\mu^a, \mu^\lambda) \in \Xi \) and a positive integer \( \hat{\alpha} \in \mathbb{Z}_{++} \) such that

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

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

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

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

\[
v^D(\alpha) = \min_{n, x, u} \{ Q(n, x, u, \alpha \mu) : (1b) - (1f), (7a), (7b) \}
\]

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

According to the monotonicity results in Corollary 1, we have

\[
Q(n, x, u, \hat{\xi}^\ell) \leq Q(n, x, u, \hat{\alpha} \mu)
\]

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

\[
\frac{1}{N} \sum_{\ell \in [N]} Q(n, x, u, \hat{\xi}^\ell) \leq Q(n, x, u, \hat{\alpha} \mu)
\]

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

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 \( v^D(\hat{\alpha}) \leq \hat{\alpha} v^D(1) \).

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 \( \alpha \lfloor q \rfloor \geq \lceil \alpha q \rceil \).

Proof: This is simply because \( \alpha \lfloor q \rfloor \geq \alpha q \) and the former is an integer. □
Suppose that \((n^*, x^*, u^*)\) is an optimal first-stage solution and \((L^1, \bar{L}^1, w^1, y^1, E^1)\) is an optimal second-stage decision to model (34) with \(\alpha = 1\).

Now let us define

\[
L_{ij}^\alpha = \min \left\{ \sum_{k \in [K]} n_{ik}^* [\delta_k c_k], \alpha L_{ij}^1 \right\}, \forall i \in [I], \forall j \in [J],
\]

\[
\bar{L}_{ij-1}^\alpha = [(1 - \mu_{ij}^\alpha) \bar{L}_{ij-1}^1], \forall i \in [I], \forall j \in [J],
\]

\[
y_{ij}^\alpha = \mathbb{I} \left( L_{ij}^\alpha < \sum_{k \in [K]} n_{ik}^* [\delta_k c_k] \right), \forall i \in [I], \forall j \in [J],
\]

\[
E_1^\alpha = \max_{i \in [I], j \in [J]} \frac{L_{ij}^\alpha}{\sum_{k \in [K]} n_{ik}^* [\delta_k c_k]},
\]

\[
E_2^\alpha = \max_{i \in [I], j \in [J]} (\alpha \mu_{ij}^\lambda)^{-1} \left( \bar{L}_{ij-1}^\alpha + \alpha \mu_{ij}^\lambda - \sum_{k \in [K]} n_{ik}^* [\delta_k c_k] \right),
\]

\[
w_{ikr}^\alpha = E_1^\alpha u_{ikr}, \forall i \in [I], \forall k \in [K], \forall r \in [R_k].
\]

According to the definition of the model (34) with \(\alpha = \alpha\), we see that \((n^*, x^*, u^*)\) is a feasible first-stage solution and \((L^\alpha, \bar{L}^\alpha, w^\alpha, y^\alpha, E^\alpha)\) 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 \(i \in [I], j \in [J]\), we must have \(\bar{L}_{ij-1}^\alpha \leq \alpha \bar{L}_{ij-1}^1\).

**Proof:** Note that

\[
\bar{L}_{ij-1}^\alpha = [(1 - \mu_{ij}^\alpha) \bar{L}_{ij-1}^1] \leq [(1 - \mu_{ij}^\alpha) \bar{L}_{ij-1}^1] \leq \alpha [(1 - \mu_{ij}^\alpha) L_{ij-1}^1] := \alpha \bar{L}_{ij-1}^1,
\]

where the first inequality is due to the definition of \(L_{ij-1}^\alpha\) and the second one is due to Claim 1. \(\diamond\)

There are two cases:

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

\[
L_{ij}^\alpha = \alpha L_{ij}^1, \bar{L}_{ij-1}^\alpha \leq \alpha \bar{L}_{ij-1}^1, L_{ij}^1 \geq \bar{L}_{ij-1}^1 + \mu_{ij}^\lambda
\]

where the first inequality is due to Claim 2 and the second one is due to the fact that \((L^1, \bar{L}^1, w^1, y^1, E^1)\) is an optimal (and of course feasible) second-stage decision to the model (34) with \(\alpha = 1\).
When \( L_{ij}^\alpha = \sum_{k \in [K]} n_{ik}^* \delta_{ck} \), we have \( y_{ij}^\alpha = 0 \). It is sufficient to show that \( \tilde{L}_{ij}^\alpha \geq \tilde{L}_{ij}^- \). This is true since

\[
\tilde{L}_{ij}^- = [(1 - \mu_{ij})L_{ij}^-] = \sum_{k \in [K]} n_{ik}^* \delta_{ck} := \tilde{L}_{ij}^a
\]

where the inequality is due to the fact that \( \tilde{L}_{ij}^a \leq \sum_{k \in [K]} n_{ik}^* \delta_{ck} \).

Q.E.D.

Finally, we observe that the objective function is

\[
E_{1i}^\alpha + \omega E_{2i}^\alpha = \max_{i \in [I], j \in [J]} \frac{\tilde{L}_{ij}^a}{\sum_{k \in [K]} n_{ik}^* \delta_{ck}} + \omega \max_{i \in [I], j \in [J]} (\tilde{\alpha} \mu_{ij}^\lambda)^{-1} \left( \tilde{L}_{i,j}^- + \tilde{\alpha} \mu_{ij}^\lambda - \sum_{k \in [K]} n_{ik}^* \delta_{ck} \right)
\]

\[
= \max_{i \in [I], j \in [J]} \frac{\min \{ \sum_{k \in [K]} n_{ik}^* \delta_{ck}, \tilde{\alpha} L_{ij}^A \} \tilde{\alpha} \min \{ \sum_{k \in [K]} n_{ik}^* \delta_{ck}, L_{ij}^A \}}{\sum_{k \in [K]} n_{ik}^* \delta_{ck}} = \tilde{\alpha} E_{1i}^D(1)
\]

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

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

A.3. Proof of Lemma 2

Lemma 2 The following characterization holds for \( \text{conv}(X_{ij,k}^2 \cap \{ x_{ik} = 1 \}) \), i.e.,

\[
\text{conv}(X_{ij,k}^2 \cap \{ x_{ik} = 1 \}) = \text{conv}(\tilde{X}_{ij,k}^2) \cap \left\{ (w_{i,:}, L_{ij}^e) : \sum_{r \in [R_k]} 2r^{-1} \delta_{ck} w_{ikr}^e \geq L_{ij}^e \right\}
\]

where

\[
\tilde{X}_{ij,k}^2 = \{(x_i, u_{i,:}, L_{ij}^e, w_{i,:}^e, E_{ij}^e) : (x_i, u_{i,:}) \in X_{ik}^e, w_{ikr}^e = u_{ikr}, E_{ij}^e, L_{ij}^e \in \mathbb{R}_+, E_{ij}^e, L_{ij}^e \in [0, 1], x_{ik} = 1 \}
\]

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

\[
\text{conv}(X_{ij,k}^2 \cap \{ x_{ik} = 1 \}) \subseteq \text{conv}(\tilde{X}_{ij,k}^2) \cap \left\{ (w_{i,:}, L_{ij}^e) : \sum_{r \in [R_k]} 2r^{-1} \delta_{ck} w_{ikr}^e \geq L_{ij}^e \right\}
\]

It remains to show that

\[
\text{conv}(X_{ij,k}^2 \cap \{ x_{ik} = 1 \}) \supseteq \text{conv}(\tilde{X}_{ij,k}^2) \cap \left\{ (w_{i,:}, L_{ij}^e) : \sum_{r \in [R_k]} 2r^{-1} \delta_{ck} w_{ikr}^e \geq L_{ij}^e \right\}
\]
Indeed, for any \((x, i, u, i, j, w, i, j, E, 1) \in \text{conv}(X_{ij}^2) \cap \left\{ (w, i, j, L, i, j) : \sum_{r \in [R_e]} 2^{r-1} [\delta, c, k] u, r, k \geq L, i, j \right\}\), there exists a finite collection \(\{(x, i, u, i, j, L, i, j, w, i, j, E, 1) \tau \in [q]\}\) and \(\{\alpha, \tau \in [q] \subset [0, 1]\}\) such that

\[
\sum_{\tau \in [q]} \alpha, \tau (x, i, u, i, j, L, i, j, w, i, j, E, 1) = (x, i, u, i, j, L, i, j, w, i, j, E, 1), \sum_{\tau \in [q]} \alpha, \tau = 1
\]

and \(\beta := \sum_{\tau \in [q]} \alpha, \tau \sum_{r \in [R_e]} 2^{r-1} [\delta, c, k] u, r, k \geq L, i, j\).

Now let us define

\[
\widehat{L, i, j} = \frac{L, i, j}{\beta} \sum_{r \in [R_e]} 2^{r-1} [\delta, c, k] u, r, k
\]

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

\[
\sum_{\tau \in [q]} \alpha, \tau (x, i, u, i, j, \widehat{L, i, j}, w, i, j, E, 1) = (x, i, u, i, j, L, i, j, w, i, j, E, 1),
\]

\[
(x, i, u, i, j, \widehat{L, i, j}, w, i, j, E, 1) \in X_{ij}^2 \cap \{x, i, k = 1\}, \forall \tau \in [q].
\]

Thus, \((x, i, u, i, j, L, i, j, w, i, j, E, 1) \in \text{conv}(X_{ij}^2 \cap \{x, i, k = 1\})\).

\[\square\]

**Appendix B: Testing Integrality of \(L, i, j\)**

We tested six small instances to compare the objective value of MILP formulations with the integer \(L, i, j\) as well as the objective value of MILP with relaxing the integrality of \(L, i, j\), where the number of bus routes \(I = 5\) and the number of scenarios \(N = 6\). 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 \(\eta = (4, 6, 3), \delta = 0.25, \omega = 5, \theta = 0, \gamma = 100\). 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 \(L, i, j\), 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 \(L, i, j\) 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 \(L, i, j\) is necessary for the sake of solution quality.
### Table 11  Objective Value of MILP with Integer $L$ and MILP with Continuous $L$

<table>
<thead>
<tr>
<th>Instance</th>
<th>Integer $L$</th>
<th>Continuous $L$</th>
<th>Relative Difference(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obj.Val</td>
<td>Time</td>
<td>Obj.Val</td>
</tr>
<tr>
<td>1</td>
<td>1.148</td>
<td>11.0</td>
<td>0.948</td>
</tr>
<tr>
<td>2</td>
<td>1.241</td>
<td>16.4</td>
<td>1.034</td>
</tr>
<tr>
<td>3</td>
<td>1.178</td>
<td>22.9</td>
<td>1.022</td>
</tr>
<tr>
<td>4</td>
<td>1.106</td>
<td>22.5</td>
<td>0.986</td>
</tr>
<tr>
<td>5</td>
<td>1.104</td>
<td>18.8</td>
<td>0.978</td>
</tr>
<tr>
<td>6</td>
<td>1.330</td>
<td>13.3</td>
<td>1.163</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Obj.Val</th>
<th>Integer $L$</th>
<th>Continuous $L$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bus 1</td>
<td>Bus 2</td>
</tr>
<tr>
<td>0.942</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>0.916</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Route</th>
<th>Scenario</th>
<th>Dominant</th>
<th>Bus 1</th>
<th>Bus 2</th>
<th>Bus 3</th>
<th>Dominant</th>
<th>Bus 1</th>
<th>Bus 2</th>
<th>Bus 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1, 3-6</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>1-6</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>