On a Fair and Risk-averse Urban Air Mobility Resource Allocation Problem Under Demand and Capacity Uncertainties

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

Haoyun Deng
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, hdeng81@gatech.edu

Peng Wei
Department of Mechanical & Aerospace Engineering, George Washington University, Washington, DC 20052, pwei@gwu.edu

Weijun Xie
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, wxie@gatech.edu

Urban Air Mobility (UAM) is an emerging air transportation mode to alleviate the ground traffic burden and achieve zero direct aviation emissions. Due to the potential economic scaling effects, the UAM traffic flow is expected to increase dramatically once implemented, and its market can be substantially large. To be prepared for the era of UAM, we study the Fair and Risk-averse Urban Air Mobility resource allocation model (FairUAM) under passenger demand and airspace capacity uncertainties for fair, safe, and efficient aircraft operations. FairUAM is a two-stage model, where the first stage is the aircraft resource allocation, and the second stage is to fairly and efficiently assign the ground and airspace delays to each aircraft provided the realization of random airspace capacities and passenger demand. We show that FairUAM is NP-hard even when there is no delay assignment decision or no aircraft allocation decision. Thus, we recast FairUAM as a mixed-integer linear program (MILP) and explore model properties and strengthen the model formulation by developing multiple families of valid inequalities. The stronger formulation allows us to develop a customized exact decomposition algorithm with both Benders and L-shaped cuts, which significantly outperforms the off-the-shelf solvers. Finally, we numerically demonstrate the effectiveness of the proposed method and draw managerial insights when applying FairUAM to a real-world network.

Key words: Urban Air Mobility, Fairness, Risk-Averse, Mixed-Integer Linear Programming, Resource Allocation

1 Introduction

The urban population in the U.S. is expected to increase from 83% in 2020 to 89% in 2050 (United Nations 2018). Albeit benefiting people with better opportunities, urbanization leads to

more severe traffic congestion problems with higher traffic volume, larger travel delays, and more energy consumption and air pollution (Schrank et al. 2021). This calls for efficient alternative transportation modes (Speranza 2018). One promising way is the on-demand point-to-point Urban Air Mobility (UAM) service, which relies on automated electric Vertical Take-Off and Landing vehicles (eVTOLs) to transport passengers at lower altitudes in and around metropolitan areas. It has been recognized recently that the UAM has the potential to improve personal mobility by reducing travel time and cost, reduce energy consumption and air pollution by using eVTOL with zero direct emission, as well as enhance economic vitality (Price et al. 2020). In the meantime, thanks to the emerging aircraft technologies in increasing reliability, aviation companies, including Joby, Archer, Eve, etc., can launch their products and prototypes to make UAM possible in the near future. Due to the potential economic scaling effects, the UAM traffic flow is expected to increase dramatically once implemented, and its market can be substantially large (Lineberger et al. 2018). To be prepared for the era of UAM, NASA recently established a UAM framework from different aspects (Price et al. 2020). One important aspect lies in UAM operations management, namely,

> *"Provide airspace operations management services as well as fleet operations management services that ensure safe, efficient, scalable, and resilient UAM operations in and around metropolitan areas."*

This high-level view echoes that a reliable and robust UAM traffic management system is required to guarantee safety and efficiency for high-density UAM traffic. Different from conventional air traffic management, UAM traffic managers and service providers are facing new challenges in cost reduction and safety promise due to higher operation density and frequency, new operating environment design, and aircraft battery capacity limitations. One possible solution is to design a sophisticated centralized traffic management system to guarantee a safe and efficient operation environment within a complex UAM network to coordinate among stakeholders, including aircraft operators, airspace managers, data providers, etc., especially when facing large-scale aircraft operations, as mentioned in Chapter 4.4.3 in Fontaine (2023).

Within a centralized UAM traffic management system, the UAM traffic manager has all the shared information from the multiple UAM service providers (see Chapter 4.3.4 in Fontaine 2023) and airspace data support from the third party, including weather and hazard forecasts *a priori* or during the flight. The UAM traffic manager is responsible for suggesting aircraft allocation to their flight paths, balancing the UAM traffic flow and airspace capacity, and taking the uncertainties into consideration. The first type of uncertainty comes from fluctuations in passenger demand (see Chapter 4.4.2 in Fontaine 2023), while the second one is the airspace capacity uncertainty that arises from stochastic weather events (see Chapter 4.6 in Fontaine 2023). Specifically, when the air traffic demand exceeds the airspace capacity, air traffic congestion occurs at some bottleneck

points, i.e., busy areas in the airspace, which causes airborne delay. Since it will be safety-critical for the eVTOL aircraft to experience long airborne delay due to its limited battery life and aircraft payload, we must manage the risk of unpredictable long airborne delay. In the UAM network, balancing the UAM traffic flow and airspace capacity in the UAM network with a proper traffic management system is of great importance to hedge against potential risks as well as to enforce safety. In this paper, we focus on managing the UAM traffic in a risk-averse way under passenger demand and airspace capacity uncertainties to minimize the airspace-level total operation cost. In addition, sorely minimizing the total cost may cause some UAM service providers to bear the most delays (Lulli and Odoni 2007). To achieve company-level fairness, we also would like to assign the airborne delay fairly among different aircraft. Therefore, this work seeks an appropriate trade-off between system-level efficiency (i.e., simultaneously minimizing the total operation cost) and achieving company-level fairness. More importantly, following the literature (Zhu et al. 2018a), it is common in the airline industry that the aircraft assignment should be available hours before the aircraft takes off due to their nontrivial travel time and safety concerns. It is also worth of mentioning that the proposed modeling framework can address other uncertainties, such as UAM service supplier uncertainty.

1.1 Literature Review

UAM literature spreads various research fields, including market study, appropriate regulation, social impact, vehicle development, and infrastructure design. Key operational constraints anticipated to impact on-demand air mobility were proposed and addressed in Vascik and Hansman (2017). For selecting appropriate vertiport locations, Daskilewicz et al. (2018) studied an integer programming model to maximize the population-cumulative potential time savings. Besides, Rajendran and Zack (2019) developed an iterative constrained clustering method to recommend potential facility locations and Rath and Chow (2022) proposed a mixed-integer linear programming (MILP) model to maximize the revenue and ridership. We refer interested readers to Garrow et al. (2021) and Rajendran and Srinivas (2020) for a comprehensive review of UAM, including its history, aircraft technology, demand prediction, infrastructure design, and operations.

Regarding the UAM air traffic management, the work of Thipphavong et al. (2018) is among the first that expanded the NASA operational concepts for UAM. Mueller et al. (2017) proposed a framework for on-demand mobility airspace, including demand-capacity balancing and trajectory planning. Kleinbekman et al. (2018) introduced an energy-efficient trajectory optimization model for UAM arrival with route selection capability, proposing a mixed-integer linear program to determine aircraft arrival sequence and minimize total arrival delay. Bharadwaj et al. (2019) presented a decentralized procedure for route planning on individual vehicles using Markov decision processes.

Bharadwaj et al. (2021) designed a scalable planning framework with safety-oriented constraints and minimized the violations. Wu and Zhang (2021) designed an integer programming model considering potential UAM traffic flow to decide the optimal vertiport locations and user allocation to vertiports. Wang et al. (2022) developed an aggregate three-dimension air traffic assignment framework to plan future UAM operations.

As a generic framework for decision-making under uncertainty (Shapiro et al. 2021), stochastic programming has been widely used in resource allocation problems in transportation for decades (Powell and Topaloglu 2003). Mukherjee and Hansen (2009) developed three stochastic air traffic flow management (ATFM) models with static ground delay and dynamic rerouting decisions. Ganji et al. (2009) proposed a stochastic optimization model to assign ground delay and optimize rerouting decisions at the same time by considering the en-route capacity. Gupta and Bertsimas (2011) presented a robust and adaptive ATFM framework under airspace capacity uncertainty, which was solved by piecewise affine policies. Chen et al. (2017) derived a chance-constrained model based on probabilistic airspace capacity and developed a polynomial approximation-based approach to solve a large-scale problem. Zhu et al. (2018b) developed a heuristic saturation technique under demand uncertainty to find the demand-independent optimal planned acceptance rates for the constrained airspace resources. Wang and Jacquillat (2020) proposed a two-stage stochastic integrated model to optimize aircraft scheduling and ground-holding operations under operating uncertainty and developed a decomposition algorithm with dual integer cuts to solve it. In volatile environments, instead of optimizing the average performance using risk-neutral objectives, risk aversion is commonly used to hedge against uncertainty and enforce safety in the worst case. Shehadeh (2022) proposed a distributionally robust optimization (DRO) model to plan fleet-sizing, routing, and scheduling for mobile facilities to minimize the conditional value at risk (CVaR) of the operational cost. Sun et al. (2022) studied a DRO fair model to optimize the public transit resource allocation during a pandemic. Moug et al. (2022) formulated a two-stage stochastic model to recruit drivers and optimize resource planning with constraints on the total number of unserved evacuees. All these works, unfortunately, do not incorporate operators' equity in their formulations, which has been regarded as one of the important features in the UAM strategic planning according to Fontaine (2023).

Besides hedging against uncertainty, fairness is also another critical issue being considered in transportation resource allocation problems. Bertsimas et al. (2011) presented a deterministic integer programming model to assign delay fairly for large-scale instances of ATFM problem. Rodionova et al. (2017) addressed fairness by adding the min-max fairness measure into their deterministic model to balance the average airline costs among different airline companies. Hamdan et al. (2018) developed an ATFM model with rerouting under deterministic capacity considering delay

distribution and flight reversal fairness among different airlines. Hamdan et al. (2022) proposed a deterministic model with central authority considering path rerouting and options and inter-flight and inter-airline fairness. Following the literature, we also apply the conventional min-max fairness and the fundamental difference between ours and the literature lies in the model formulations since the UAM often involves short-distance flights and eVTOLs, where addressing uncertainties is critical to its safety.

Particularly, similar to many airline resource allocation problems, within a centralized UAM traffic management system (e.g., under the guideline of FAA), UAM service providers are collaboratively managing the resource allocation to achieve systematic efficiency and reduce risk at the same time. On the other hand, its decision should also address the equity among different participating service providers. In fact, fairness and risk in UAM operations are more critical due to aircraft limitations (Connors 2020). However, the literature on fair or risk-averse UAM traffic management is rather limited. Pelegrín et al. (2021) proposed a deconfliction model to minimize the total deviation from flight schedules in a fair way and Wu et al. (2022) introduced a path planning algorithm for multiple aircraft with fairness awareness to generate collision-free paths under the vehicle and environmental uncertainties. Chin et al. (2021) focused on three fairness metrics for on-demand air mobility resource allocation in a rolling horizon to incorporate the dynamic demand. Bennaceur et al. (2022) proposed a MILP model for the UAM air-taxi problem with fairness between the regular and premium classes. To efficiently solve the problem, they decomposed the original model into two subproblems, then used a beam-search algorithm to solve the "Pooling and Scheduling" part and a variable neighborhood search algorithm to solve the "Routing and Recharging" part. To the best of our knowledge, Hou et al. (2021) is the only work on risk-averse stochastic UAM network design in a data-driven framework, considering vertiport location selection and aircraft route planning. Different from Hou et al. (2021), we optimize disaggregate traffic management not only under passenger demand uncertainty but also under airspace capacity uncertainty related to weather events. Besides, we incorporate alternative routes on each origin-destination (OD) pair and intersections of different routes in the airspace. As far as we are concerned, our model is the first one focusing on fair and risk-averse UAM resource allocation under passenger demand and airspace capacity uncertainties.

1.2 Summary of Main Contributions

This paper studies the Fair and Risk-averse Urban Air Mobility resource allocation model (FairUAM) to coordinate multiple UAM service providers. The goal of FairUAM is to determine the optimal UAM aircraft resource allocation and delay assignment under passenger demand and airspace uncertainties to minimize the weighted sum of total cost and highest average company

cost to achieve both system-level efficiency (i.e., total operation cost) and company-level fairness. The main contributions of FairUAM are summarized below.

i. We develop a novel formulation for FairUAM to allocate aircraft from multiple UAM service providers under passenger demand and airspace capacity uncertainties. The proposed FairUAM is versatile and can handle the unsatisfied demand (i.e., the UAM traffic flow needed exceeds the airspace capacity when passenger demand is very large) and can also address company-level fairness as well as achieve the overall airspace efficiency.

ii. Without considering delay assignment, we prove that FairUAM reduces to UAM aircraft resource allocation problem which is NP-hard even with one UAM service provider. Given aircraft allocation, FairUAM reduces to UAM delay assignment problem, which is also NP-hard. These complexity results motivate us to study an effective solution algorithm to solve FairUAM.

iii. We derive monotonicity properties such that an increase in passenger demand, a decrease in airspace capacity or an increase in unsatisfied passenger demand leads to a possible increase in objective value. While achieving the same optimality as the original model, these properties allow us to simplify the FairUAM model and further relax the integrality of some decision variables for computational advantages.

iv. To strengthen the formulation, we develop several families of valid inequalities by exploring the model structures and discovering the hidden relationship between different decision variables. Specifically, we derive airspace guarantee inequalities, symmetry-breaking inequalities, data pre-processing inequalities, bottleneck point arrival time inequalities, total arrival time inequalities, and accumulated delay inequalities. These valid inequalities allow us to design and strengthen the decomposition-based algorithm, which can solve larger-scale instances with 5 times more binary variables compared to solving the vanilla model directly via the off-the-shelf solvers.

v. We numerically demonstrate the effectiveness of the proposed method using generated instances and apply it to a case study of a real-world network. We find that with only 20% of increase in the total operation cost, we can reduce the unfairness by 68%. We also observe that the risk-averse solution is more robust compared to the risk-neutral one, when accurate estimation of the uncertainties is not practical. We show that our framework can be useful in identifying the network hotspots and supporting UAM service providers with better path-planning insights.

Notation For readers' convenience, we list the notation used throughout the paper. Bold letters (e.g., $\boldsymbol{y}$) are used to represent vectors and matrices and the corresponding non-bold letters denote their components. We let bold script letters (e.g., $\mathcal{R}$) denote sets and $|\mathcal{R}|$ to denote their cardinality.

We use $\widetilde{\boldsymbol{\xi}}$ to represent a random vector and let $\boldsymbol{\xi}$ be its realizations. Additional notations related to our model will be introduced in the next section.

Organization The remainder of the paper is organized as follows. In Section 2, we introduce the problem setting and model formulation. Section 3 derives model properties. Section 4 proposes valid inequalities and develops a decomposition algorithm. Finally, Section 5 numerically demonstrates the performance of the proposed valid inequalities and algorithm, and Section 6 concludes this paper.

2 Model Formulation

2.1 Overview

In this work, we consider a network, as shown in Figure 1, which consists of multiple UAM service providers each having their own aircraft, and multiple origin-destination (OD) pairs each having multiple candidate routes, under airspace capacity and passenger demand uncertainties. Typically, the intersections of routes are considered as bottleneck points where congestion may occur. Besides, weather events at bottleneck points further influence their capability to accommodate air traffic flow, which leads to airspace capacity uncertainty at bottleneck points. As on-demand services, passengers are allowed to make the reservation including an origin and a destination, a day or hours (i.e., our planning horizon is a day or half-day) before the actual departure time, which in turn brings in passenger demand uncertainty. In our framework, the UAM traffic manager optimizes the operation plan for the whole planning horizon (i.e., before the realization of uncertainties) by assigning the aircraft to OD pairs and routes. After the UAM service providers submit their realized passenger demand, the UAM traffic manager assigns delays to aircraft to achieve efficiency and company-level fairness at the same time.

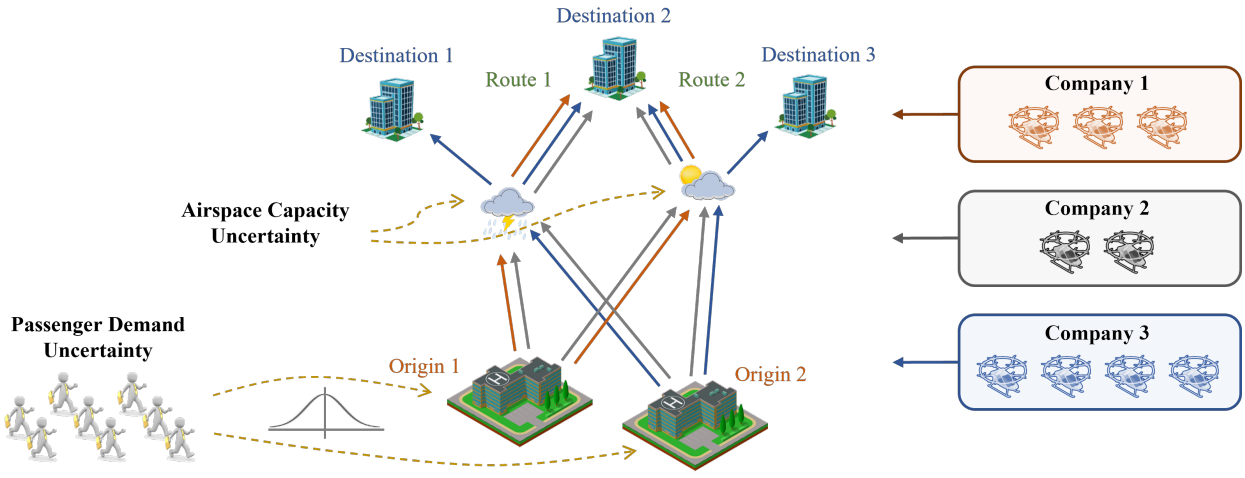

Figure 1 An illustration of the FairUAM network and its components

The FairUAM consists of a set of UAM service providers $\mathcal{L}$, where the service provider $\ell \in \mathcal{L}$ owns $|\mathcal{I}_\ell|$ aircraft. Each service provider (or company) would like to satisfy their own passenger demand to transport passengers from origin vertiports to destination vertiports in their favor. Each aircraft can transport a certain number of passengers with the same origin and destination. We let m_i be the passenger capacity for each aircraft $i \in \mathcal{I}_\ell$ for each $\ell \in \mathcal{L}$ and $\mathcal{R}$ represent the set of OD pairs. While multiple aircraft are operated on different OD pairs, intersections between OD pairs may exist and congestion may happen. To circumvent aircraft conflict on such intersected OD pairs, we suppose that the airspace network consists of a set of bottleneck points $\mathcal{K}$ for flow management. There are multiple routes for each aircraft to choose from, where for a given OD pair, some aircraft need to reroute or detour such that the number of aircraft passing the same bottleneck within a time unit is restricted within its capacity. Particularly, we let $\mathcal{J}_r$ denote the sets of routes for each OD pair $r \in \mathcal{R}$. To generate a better operation plan for upcoming travel needs, we consider a planning horizon $\mathcal{T}$ which includes finite time units. The length of a time unit can be changed based on our needs, for example, according to aircraft density and different operation requirements.

Since UAM operations are quite sensitive to severe weather events, airspace capacity is random due to unexpected weather conditions and safety. Particularly, we let a random parameter $\widetilde{C}_t^k \in \mathbb{Z}_+$ denote the airspace capacity of each bottleneck point $k \in \mathcal{K}$ at a given time unit $t \in \mathcal{T}$. The objective of FairUAM is to dispatch aircraft to OD pairs and routes based on companies' needs, but it may delay some aircraft to achieve better utilization of the limited airspace capacities and minimize the operation cost. As an on-demand service, passenger demand is stochastic since passengers are allowed to book their seats just one day or even hours before their departure time. Considering the relocation of aircraft, passenger demand may be realized after the aircraft assignment. We let the random parameter $\widetilde{D}_\ell^r \in \mathbb{Z}_+$ denote the passenger demand of service provider $\ell \in \mathcal{L}$ on OD pair $r \in \mathcal{R}$. For notational convenience, we let random parameters $\widetilde{\boldsymbol{\xi}} = (\widetilde{\boldsymbol{C}}, \widetilde{\boldsymbol{D}})$. Note that the UAM traffic manager and service providers make decisions independently, following all the regulations in an information-sharing environment. This may not necessarily lead to efficient aircraft operations and airspace utilization. In this context, the decisions in FairUAM are made by a central authority on proposing a collaborative UAM air traffic management system.

2.2 Model Decisions

In our model, the operation cost for each service provider $\ell \in \mathcal{L}$ includes assignment cost, relocation cost, delay cost, and penalty cost. The assignment cost is the fixed travel cost for an aircraft assigned to a route, including the cost needed in the free-condition airspace, denoted by c_{ij}^r for aircraft $i \in \mathcal{I}_\ell$ assigned to OD pair $r \in \mathcal{R}$ and route $j \in \mathcal{J}_r$. The relocation cost is the fixed cost for

an aircraft assigned to an OD pair, denoted by q_i^r for aircraft $i \in \mathcal{I}_\ell$ relocated to OD pair $r \in \mathcal{R}$. The delay cost includes ground delay cost α_{ij}^r at the origin vertiport and airborne delay cost β_{ij}^r at each bottleneck point for aircraft $i \in \mathcal{I}_\ell$ assigned to OD pair $r \in \mathcal{R}$ and route $j \in \mathcal{J}_r$. An UAM service provider ℓ bears a penalty cost p_ℓ^r for the business loss when they are not able to satisfy its passenger demand on OD pair $r \in \mathcal{R}$. Besides minimizing the total operation cost, we also consider inter-airline fairness, i.e., the average operation cost per aircraft for each airline. In this case, we also include the fairness term into our objective so that the largest average airline operation cost is also minimized. To better balance the two goals, we introduce a weighted scalar λ to balance efficiency and fairness. Therefore, the objective of FairUAM is to minimize the weighted sum of the total operation costs and the largest average airline operation cost.

The traffic manager in FairUAM needs to decide OD pair and route assignment for each aircraft upfront and determine the aircraft's delays after random parameters such as passenger demand and airspace capacity are realized. We let $y_i^r \in \{0,1\}$ denote whether aircraft $i \in \mathcal{I}_\ell$ of service provider $\ell \in \mathcal{L}$ is assigned to OD pair $r \in \mathcal{R}$ and let $\delta_{ij}^r \in \{0,1\}$ denote whether it is assigned to route $j \in \mathcal{J}_r$ on OD pair $r \in \mathcal{R}$ or not. In FairUAM, given a realization $\boldsymbol{\xi}$ of the random parameters $\widetilde{\boldsymbol{\xi}}$, we let $g_{ij}^r(\boldsymbol{\xi})$ denote the ground delay of aircraft $i \in \mathcal{I}_\ell$ of service provider $\ell \in \mathcal{L}$ on route $j \in \mathcal{J}_r$ and OD pair $r \in \mathcal{R}$ and let the corresponding airborne delay at bottleneck point $k \in \mathcal{K}$ be $a_{ijk}^r(\boldsymbol{\xi})$. When the number of aircraft required to satisfy the passenger demand exceeds the available airspace capacity or the number of available aircraft, some passengers may be turned down, and the service provider will incur a penalty cost for the loss of business. To model this effect, we let $s_\ell^r(\boldsymbol{\xi})$ denote the unsatisfied passenger demand of service provider $\ell \in \mathcal{L}$ on OD pair $r \in \mathcal{R}$.

2.3 The Risk-averse First-stage Model Formulation

To better ensure safety, it is important to hedge against the risk of passenger demand and airspace capacity uncertainties in the UAM operating environment. For FairUAM, we consider the Conditional Value-at-Risk (CVaR) (Rockafellar et al. 2000),

$$\rho(Z) = \mathrm{CVaR}_{1-\varepsilon}(Z) := \inf_{v \in \mathbb{R}} \left\{ v + \frac{1}{\varepsilon} \mathbb{E}[(Z-v)^+] \right\}$$

where $(n)^+ := \max\{n, 0\}$ and $v = \mathrm{VaR}_{1-\varepsilon}(Z) := \inf\{v : \mathbb{P}(Z \le v) \ge 1-\varepsilon\}$. We can simply extend our model to other coherent risk measures. For simplicity and generality, we mainly focus on the most representative one- CVaR, which has been widely applied to transportation literature (Faghih-Roohi et al. 2016, Toumazis and Kwon 2013, Lei et al. 2018, Yu et al. 2021a). Note that CVaR naturally bridges the conventional stochastic program (i.e., by setting $\epsilon = 1$) and the robust optimization (i.e., by setting $\epsilon = 0$). Besides, CVaR can be interpreted as a distributionally robust

optimization with interval ambiguity set (Shapiro and Ahmed 2004, Jiang and Xie 2023). Under this setting, FairUAM is given by:

$$\min_{\boldsymbol{\delta},\boldsymbol{y}} \quad \rho\left[\mathcal{Q}\left(\boldsymbol{y},\boldsymbol{\delta},\widetilde{\boldsymbol{\xi}}\right)\right], \tag{1a}$$

$$\text{s.t.} \quad \sum_{r\in\mathcal{R}} y_i^r \le 1, \quad \forall i\in\mathcal{I}_\ell, \ell\in\mathcal{L}, \tag{1b}$$

$$\sum_{j\in\mathcal{J}_r} \delta_{ij}^r = y_i^r, \quad \forall r\in\mathcal{R}, i\in\mathcal{I}_\ell, \ell\in\mathcal{L}, \tag{1c}$$

$$y_i^r, \delta_{ij}^r \in \{0,1\}, \quad \forall j\in\mathcal{J}_r, r\in\mathcal{R}, i\in\mathcal{I}_\ell, \ell\in\mathcal{L}. \tag{1d}$$

In FairUAM (1), the objective is to minimize the risk of the planning outcome and $\mathcal{Q}(\boldsymbol{y},\boldsymbol{\delta},\widetilde{\boldsymbol{\xi}})$ denotes the random recourse function, which is specified in the next paragraph. Constraints (1b) guarantee that each aircraft can only be assigned to at most one OD pair. Constraints (1c) ensure that if an aircraft is assigned to an OD pair, it can be assigned to only one route. Otherwise, it will not be assigned to any route. Constraints (1d) specify the boundaries of the decision variables. Note that in (1), the variables $\boldsymbol{y}$ may not be necessary. However, we keep them for the natational convenience when introducing valid inequalities.

2.4 The Second-stage Model Formulation

Given a realization $\boldsymbol{\xi}$ of random parameters $\widetilde{\boldsymbol{\xi}}$ and the values of first-stage decisions $(\boldsymbol{y},\boldsymbol{\delta})$, we can formulate the recourse function as follows:

$$\mathcal{Q}(\boldsymbol{y},\boldsymbol{\delta},\boldsymbol{\xi}) =$$

$$\min_{\boldsymbol{g}(\boldsymbol{\xi}),\boldsymbol{a}(\boldsymbol{\xi}),\boldsymbol{s}(\boldsymbol{\xi}),\boldsymbol{B}(\boldsymbol{\xi})} Q(\boldsymbol{y},\boldsymbol{\delta},\boldsymbol{\xi},\boldsymbol{g}(\boldsymbol{\xi}),\boldsymbol{a}(\boldsymbol{\xi}),\boldsymbol{s}(\boldsymbol{\xi}),\boldsymbol{B}(\boldsymbol{\xi})) := \left\{\lambda\sum_{\ell\in\mathcal{L}}\frac{\phi_\ell(\boldsymbol{\xi})}{|\mathcal{I}_\ell|} + (1-\lambda)\max_{\ell\in\mathcal{L}}\frac{\phi_\ell(\boldsymbol{\xi})}{|\mathcal{I}_\ell|}\right\}, \tag{2a}$$

$$\text{s.t.} \quad \phi_\ell(\boldsymbol{\xi}) = \sum_{r\in\mathcal{R}}\left\{p_\ell^r s_\ell^r(\boldsymbol{\xi}) + \sum_{i\in\mathcal{I}_\ell}\left[q_i^r y_i^r + \sum_{j\in\mathcal{J}_r}\left(c_{ij}^r\delta_{ij}^r + \alpha_{ij}^r g_{ij}^r(\boldsymbol{\xi}) + \beta_{ij}^r\sum_{k\in\mathcal{K}} a_{ijk}^r(\boldsymbol{\xi})\right)\right]\right\}, \forall \ell\in\mathcal{L}, \tag{2b}$$

$$s_\ell^r(\boldsymbol{\xi}) + \sum_{i\in\mathcal{I}_\ell} m_i y_i^r \ge D_\ell^r, \quad \forall r\in\mathcal{R}, \ell\in\mathcal{L}, \tag{2c}$$

$$g_{ij}^r(\boldsymbol{\xi}) + \sum_{k\in\mathcal{K}} a_{ijk}^r(\boldsymbol{\xi}) \le \delta_{ij}^r\bar{L}, \quad \forall j\in\mathcal{J}_r, r\in\mathcal{R}, i\in\mathcal{I}_\ell, \ell\in\mathcal{L}, \tag{2d}$$

$$\sum_{k\in\mathcal{K}} a_{ijk}^r(\boldsymbol{\xi}) \le \delta_{ij}^r\bar{R}_{ij}^r, \quad \forall j\in\mathcal{J}_r, r\in\mathcal{R}, i\in\mathcal{I}_\ell, \ell\in\mathcal{L}, \tag{2e}$$

$$\sum_{t\in\mathcal{T}} B_{it}^k(\boldsymbol{\xi}) \le 1, \quad \forall i\in\mathcal{I}_\ell, \ell\in\mathcal{L}, k\in\mathcal{K}, \tag{2f}$$

$$\sum_{t\in\mathcal{T}} tB_{it}^k(\boldsymbol{\xi}) = \sum_{r\in\mathcal{R}}\left(\sum_{j\in\mathcal{J}_r}\left(\tau_{ijk}^r\delta_{ij}^r + g_{ij}^r(\boldsymbol{\xi}) + \sum_{\substack{id_j(k)\ge id_j(k')\\ k'\in\mathcal{K}}} a_{ijk'}^r(\boldsymbol{\xi})\right)\right), \forall i\in\mathcal{I}_\ell, \ell\in\mathcal{L}, k\in\mathcal{K}, \tag{2g}$$

$$\sum_{\ell\in\mathcal{L}}\sum_{i\in\mathcal{I}_\ell} B_{it}^k(\boldsymbol{\xi}) \le C_t^k, \quad \forall k\in\mathcal{K}, t\in\mathcal{T}\setminus\{0\}, \tag{2h}$$

$$B_{it}^k(\boldsymbol{\xi}) \in \{0,1\}, \quad \forall i\in\mathcal{I}_\ell, \ell\in\mathcal{L}, k\in\mathcal{K}, t\in\mathcal{T}, \tag{2i}$$

$$g_{ij}^r(\boldsymbol{\xi}), a_{ijk}^r(\boldsymbol{\xi}), s_\ell^r(\boldsymbol{\xi}) \in \mathbb{Z}_+, \quad \forall j\in\mathcal{J}_r, r\in\mathcal{R}, i\in\mathcal{I}_\ell, \ell\in\mathcal{L}, k\in\mathcal{K}, \tag{2j}$$

$$\phi_\ell(\boldsymbol{\xi}) \ge 0, \quad \forall \ell\in\mathcal{L}. \tag{2k}$$

Here, $\boldsymbol{B}(\boldsymbol{\xi})$ are auxiliary variables enforcing the connection between time and location for each aircraft. Specifically, we let $B_{it}^k(\boldsymbol{\xi})$ denote whether aircraft $i\in\mathcal{I}_\ell$ for service provider $\ell\in\mathcal{L}$ arrives at a bottleneck point $k\in\mathcal{K}$ at time $t\in\mathcal{T}$. Objective (2a) is to minimize the weighted sum of the efficiency term (i.e., total operation cost) and the fairness term (i.e., the largest average company operation cost). Note that one may also bound the fairness term from the above. In this paper, we use the weighted sum instead of modeling the fairness term into a constraint, which is because the fairness constraint may be infeasible if we choose an inappropriate upper bound. To make sure the efficiency term is of the same magnitude as the fairness term, instead of taking the summation of the company operation costs, we use the summation of the average company operation cost. Inspired by Rodionova et al. (2017), we use the widely used min-max fairness for the fairness term (Radunovic and Le Boudec 2007, Du et al. 2017). Constraints (2b) compute the total operation cost for each service provider. Constraints (2c) ensure that the satisfied and unsatisfied passenger demand should meet the requirement. Constraints (2d) and (2e) specify the total delay and airborne delay limit and postulate the number of passengers on board at each stop, respectively. Constraints (2f) and (2g) define the auxiliary variables $\boldsymbol{B}(\boldsymbol{\xi})$ such that each aircraft can only pass a specific bottleneck point once, and the time it arrives at the bottleneck should be equal to the sum of the travel time τ_{ijk}^r needed for aircraft $i\in\mathcal{I}_\ell$ from the service provider $\ell\in\mathcal{L}$ on the OD pair $r\in\mathcal{R}$ and the route $j\in\mathcal{J}_r$ in the free-condition airspace and the accumulated ground and airborne delay before visiting this bottleneck point. Note that we use $id_j(k)$ to denote the order of bottleneck point $k\in\mathcal{K}$ being visited on route $j\in\mathcal{J}_r$ of OD pair $r\in\mathcal{R}$ and let $\sum_{id_j(k)\ge id_j(k'),k'\in\mathcal{K}} a_{ijk'}^r(\boldsymbol{\xi})$ denote the total airborne delay before visiting bottleneck point k. Constraints (2h) guarantee that the total number of aircraft at the bottleneck point within a unit of time cannot exceed the airspace capacity. Constraints (2i) - (2k) specify the boundary conditions of the decision variables and the auxiliary variables.

3 Model Properties

In this section, we first prove the complexity of FairUAM (1) and then explore its properties for formulation simplification and managerial insights.

3.1 Computational Complexity of FairUAM (1)

We observe that FairUAM (1) is a mixed-integer nonlinear program which is expected to be a difficult problem. This motivates us to develop model properties and a customized algorithm to accelerate the solution procedure. In fact, we prove that solving FairUAM (1) is NP-hard with the reduction to the well-known NP-complete partition problem.

PROPOSITION 1. *Solving FairUAM (1), in general, is NP-hard even when* $\lambda = 1, |\mathcal{L}| = 1, |\mathcal{R}| = 2, |\mathcal{J}_r| = 1$, *and the random variables* $\widetilde{\boldsymbol{\xi}}$ *have only one realization.*

Proof. See Appendix A.1. □

The complexity of Proposition 1 is owing to the binary variables in the first-stage problem. In fact, we prove that even with a given first-stage solution, the resulting problem, due to the special network structure in FairUAM (1) (i.e., bottleneck points), is still NP-hard to solve. More specifically, assigning ground and airborne delays with a given OD pair and route assignment can be an NP-hard combinatorial optimization problem. The complexity result is summarized below.

PROPOSITION 2. *Solving the second-stage problem (2) is NP-hard even when a first stage decision is given, the random parameters have one realization, and* $\lambda = 1, |\mathcal{L}| = 1, |\mathcal{J}_r| = 1$ *for all OD pair* $r \in \mathcal{R}$.

Proof. See Appendix A.2. □

Proposition 1 and Proposition 2 show that even under simple settings, FairUAM (1) may not be polynomial-time solvable. This motivates us to strengthen the original formulation and develop an exact method to solve it in the next section.

3.2 Sensitivity Analyses

In this subsection, we analyze the properties of the recourse function $\mathcal{Q}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\xi})$ and the second-stage objective function $Q(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{g}(\boldsymbol{\xi}), \boldsymbol{a}(\boldsymbol{\xi}), \boldsymbol{s}(\boldsymbol{\xi}), \boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{B}(\boldsymbol{\xi}))$. These properties allow us to further simplify FairUAM (1) in the following sections. We first introduce the results for the second-stage objective function $Q(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{g}(\boldsymbol{\xi}), \boldsymbol{a}(\boldsymbol{\xi}), \boldsymbol{s}(\boldsymbol{\xi}), \boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{B}(\boldsymbol{\xi}))$.

PROPOSITION 3. *The following hold for the second-stage objective function* $Q(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{g}(\boldsymbol{\xi}), \boldsymbol{a}(\boldsymbol{\xi}), \boldsymbol{s}(\boldsymbol{\xi}), \boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{B}(\boldsymbol{\xi}))$*:*

- *Function* $Q(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{g}(\boldsymbol{\xi}), \boldsymbol{a}(\boldsymbol{\xi}), \boldsymbol{s}(\boldsymbol{\xi}), \boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{B}(\boldsymbol{\xi}))$ *is monotone non-decreasing as the second stage decision unsatisfied trafic demand* $s_\ell^r(\boldsymbol{\xi})$ *non-decreases for some* $\ell \in \mathcal{L}, r \in \mathcal{R}$*;*
- *Function* $Q(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{g}(\boldsymbol{\xi}), \boldsymbol{a}(\boldsymbol{\xi}), \boldsymbol{s}(\boldsymbol{\xi}), \boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{B}(\boldsymbol{\xi}))$ *is monotone non-decreasing as the passenger demand* D_ℓ^r *increases for some* $\ell \in \mathcal{L}, r \in \mathcal{R}$*; and*

• *Function* $Q(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{g}(\boldsymbol{\xi}), \boldsymbol{a}(\boldsymbol{\xi}), \boldsymbol{s}(\boldsymbol{\xi}), \boldsymbol{u}(\boldsymbol{\xi}), \boldsymbol{B}(\boldsymbol{\xi}))$ *is monotone non-increasing as the airspace capacity* C_t^k *increases for some* $t \in \mathcal{T}, k \in \mathcal{K}$.

Proof. We prove the results according to their order.

• The first monotonicity result is because an increase in unsatisfied passenger demand leads to an increase in airline operation cost, i.e., both efficiency and fairness costs increase or stay the same.

• The second monotonicity result is because when the passenger demand increases, more aircraft are needed to satisfy the passenger demand which leads to an increase in airline operation cost. Thus, both efficiency and fairness costs in the objective decrease or stay the same.

• The third monotonicity result is because when the airspace capacity increases, more aircraft are allowed to pass the same bottleneck point at the same time instead of waiting before entering this bottleneck point. In addition, more aircraft if available can be used to satisfy the passenger demand which helps reduce the airline operation cost. Thus, both efficiency and fairness costs in the objective decrease or stay the same. □

As the minimization operator preserves the monotonicity of a function, the second and third results of Proposition 3 still hold for the recourse function $\mathcal{Q}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\xi})$.

Corollary 1. *The following hold for the recourse function* $\mathcal{Q}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\xi})$*:*

• *Recourse function* $\mathcal{Q}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\xi})$ *is monotone non-decreasing as the passenger demand* D_ℓ^r *increases for some* $\ell \in \mathcal{L}, r \in \mathcal{R}$*; and*

• *Recourse function* $Q(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\xi})$ *is monotone non-increasing as the airspace capacity* C_t^k *increases for some* $t \in \mathcal{T}, k \in \mathcal{K}$.

Proposition 3 and Corollary 1 motivate us that FairUAM (1) can be further simplified based on the monotonicity. It also shows that when the airspace capacity is large enough, aircraft can be assigned to OD pairs and routes without any airspace capacity limitation such that the ground delay and airborne delay can be nearly zero. In this case, when all aircraft can travel in the free-condition airspace, the original problem can be reduced to an aircraft allocation problem.

3.3 Model Simplification and Integrality of Decision Variables

In this subsection, we first simplify FairUAM (1) by linearizing the nonlinear terms, then prove that at optimality, several classes of second-stage integer decision variables can be relaxed to be continuous, which can significantly relieve the computational effort.

Linearization As mentioned at the beginning of this section, one difficulty of FairUAM (1) resides in the nonlinear terms in the first-stage objective (1a) and second-stage objective (2a). By introducing one more variable and two extra constraints, we can equivalently represent FairUAM (1) as follows:

$$\begin{aligned} \min_{\boldsymbol{\delta},\boldsymbol{y},v} \quad & v+\varepsilon^{-1}\mathbb{E}_{\mathbb{P}}\left[h(\widetilde{\boldsymbol{\xi}})\right], && \text{(3a)}\\ \text{s.t.} \quad & h(\boldsymbol{\xi}) \geq \mathcal{Q}(\boldsymbol{\delta},\boldsymbol{y},\boldsymbol{\xi}) - v, && \text{(3b)}\\ & h(\boldsymbol{\xi}), v \geq 0, && \text{(3c)}\\ & \text{(1b)} - \text{(1d)}. \end{aligned}$$

For the second-stage problem (2), we let $\psi(\boldsymbol{\xi})$ denote the fairness term. Then the second-stage problem (2) is equivalent to

$$\begin{aligned} \min_{\boldsymbol{g}(\boldsymbol{\xi}),\boldsymbol{a}(\boldsymbol{\xi}),\boldsymbol{s}(\boldsymbol{\xi}),\boldsymbol{B}(\boldsymbol{\xi})} Q(\boldsymbol{y},\boldsymbol{\delta},\boldsymbol{\xi},\boldsymbol{g}(\boldsymbol{\xi}),\boldsymbol{a}(\boldsymbol{\xi}),\boldsymbol{s}(\boldsymbol{\xi}),\boldsymbol{B}(\boldsymbol{\xi})) = & \lambda\sum_{\ell\in\mathcal{L}}\frac{\phi_\ell(\boldsymbol{\xi})}{|\mathcal{I}_\ell|} + (1-\lambda)\psi(\boldsymbol{\xi}), && \text{(4a)}\\ \text{s.t.} \quad & \phi_\ell(\boldsymbol{\xi}) - \psi(\boldsymbol{\xi})|\mathcal{I}_\ell| \leq 0, \quad \forall \ell\in\mathcal{L}, && \text{(4b)}\\ & \psi(\boldsymbol{\xi}) \geq 0, && \text{(4c)}\\ & \text{(2b)} - \text{(2k)}. \end{aligned}$$

Relaxing Integrality of Decision Variables Based on the monotonicity results and problem structure, we prove that at optimality, we can relax the integrality of variables, which can enhance the efficacy to solve FairUAM (1) more efficiently.

PROPOSITION 4. *Relaxing the integrality of $\boldsymbol{s}(\boldsymbol{\xi}),\boldsymbol{g}(\boldsymbol{\xi}),\boldsymbol{a}(\boldsymbol{\xi})$ in FairUAM* (1) *preserves the same optimal value.*

Proof. Notice that in constraints (2c) and (2d), at optimality, we must have $s^r_\ell(\boldsymbol{\xi}) = \max\{0, D^r_\ell - \sum_{i\in\mathcal{I}_\ell} m_i y^r_i\}$ for all $r\in\mathcal{R}, \ell\in\mathcal{L}$. Since D^r_ℓ, m_i, y^r_i are all integers, we can relax the integrality of variables $\boldsymbol{s}$.

For the constraint matrix related with variables $\boldsymbol{g}(\boldsymbol{\xi})$ and $\boldsymbol{a}(\boldsymbol{\xi})$, let $(\boldsymbol{y},\boldsymbol{\delta},\boldsymbol{B},\boldsymbol{g},\boldsymbol{a},\boldsymbol{s},\boldsymbol{u})$ be a feasible solution of the relaxed problem. According to (1b), (1c) and (2d), for any aircraft $i\in\mathcal{I}_\ell$, we must have $\delta^{\bar{r}}_{i\bar{j}} = 1$ if there exists $g^{\bar{r}}_{i\bar{j}} > 0$ or $a^{\bar{r}}_{i\bar{j}k} > 0$ for some $\bar{j}\in\mathcal{J}_{\bar{r}}$ and some $\bar{r}\in\mathcal{R}$, and $g^{\bar{r}}_{i\bar{j}} = a^{\bar{r}}_{i\bar{j}k} = 0$ for any $j\neq\bar{j}, \forall r\neq\bar{r}$. Let $\{k_1,k_2,\ldots,k_n\}$ be the bottleneck points in the visiting order on route $\bar{j}$ of OD pair $\bar{r}$. According to (2d), (2e), (2g), we must have

$$\begin{aligned} \delta^{\bar{r}}_{i\bar{j}}\bar{L} &\geq g^{\bar{r}}_{i\bar{j}}(\boldsymbol{\xi}) + \sum_{k\in\{k_1,\ldots,k_n\}} a^{\bar{r}}_{i\bar{j}k},\\ \delta^{\bar{r}}_{i\bar{j}}\bar{R}^{\bar{r}}_{i\bar{j}} &\geq \sum_{k\in\{k_1,\ldots,k_n\}} a^{\bar{r}}_{i\bar{j}k}, \end{aligned}$$

$$\sum_{t\in\mathcal{T}} tB^{k_1}_{i,t}(\boldsymbol{\xi}) = \tau^{\bar{od}}_{i\bar{i}\bar{j}k_1}\delta^{\bar{r}}_{i\bar{i}\bar{j}} + g^{\bar{r}}_{i\bar{i}\bar{j}} + a^{\bar{r}}_{i\bar{i}\bar{j}k_1},$$
$$\sum_{t\in\mathcal{T}} tB^{k_2}_{i,t}(\boldsymbol{\xi}) = \tau^{\bar{od}}_{i\bar{i}\bar{j}k_1}\delta^{\bar{r}}_{i\bar{i}\bar{j}} + g^{\bar{r}}_{i\bar{i}\bar{j}} + a^{\bar{r}}_{i\bar{i}\bar{j}k_1} + a^{\bar{r}}_{i\bar{i}\bar{j}k_2},$$
$$\vdots$$
$$\sum_{t\in\mathcal{T}} tB^{k_n}_{i,t}(\boldsymbol{\xi}) = \tau^{\bar{r}}_{i\bar{i}\bar{j}k_1}\delta^{\bar{r}}_{i\bar{i}\bar{j}} + g^{\bar{r}}_{i\bar{i}\bar{j}} + \sum_{k\in\{k_1,\ldots,k_n\}} a^{\bar{r}}_{i\bar{i}\bar{j}k},$$

for each $i \in \mathcal{I}$. We see that for each $i \in \mathcal{I}$, the constraint matrix corresponding to variables $\boldsymbol{g}(\boldsymbol{\xi})$ and $\boldsymbol{a}(\boldsymbol{\xi})$ is an interval matrix. Thus, the overall constraint matrix corresponding to variables $\boldsymbol{g}(\boldsymbol{\xi})$ and $\boldsymbol{a}(\boldsymbol{\xi})$ is an interval matrix. Hence, according to corollary 3 in Kong et al. (2013), the overall constraint matrix related to variables $\boldsymbol{g}(\boldsymbol{\xi})$ and $\boldsymbol{a}(\boldsymbol{\xi})$ is totally unimodular. Thus, both variables $\boldsymbol{g}(\boldsymbol{\xi}), \boldsymbol{a}(\boldsymbol{\xi})$ must be integral when fixing the other variables. □

Since in most practical situations, an airborne delay is much more expensive than a ground delay due to the limitation in aircraft and airspace, Proposition 4 can improve the second-stage problem (2). We also demonstrate the effectiveness of Proposition 4 in Section 5.1.

4 Valid Inequalities and A Solution Method

In this section, we explore the formulation structure and strengthen FairUAM (1) by developing different families of valid inequalities and based on these efforts, we develop an efficient decomposition algorithm.

4.1 Valid Inequalities

In this subsection, we develop different families of valid inequalities to strengthen FairUAM (1). All the inequalities are valid for the entire mixed-integer decisions of FairUAM (1) except the symmetry-breaking ones, where the latter cut off redundant alternative mixed-integer decisions. All the valid inequalities are obtained by investigating the model properties.

Airspace Guarantee Inequalities According to constraints (1c), if an aircraft is assigned to an OD pair, it must be assigned to a particular route corresponding to the OD pair. Consequently, the aircraft must go through the bottleneck points on its assigned route at some time points during the planning horizon according to constraints (2f) and (2g). Otherwise, no bottleneck point will be visited when the aircraft is idling. Let ω is the minimum number of bottleneck points that must be visited among all the possible OD pairs. Then, if $\sum_{r\in\mathcal{R}} y^r_i = 1$, we must have $\omega \le \sum_{t\in\mathcal{T}} B^k_{it}(\boldsymbol{\xi}) \le 1$. This motivates us to derive the following valid inequalities:

$$\sum_{t\in\mathcal{T}} B^k_{it}(\boldsymbol{\xi}) \le \sum_{r\in\mathcal{R}} y^r_i, \quad \forall i \in \mathcal{I}_\ell, \ell \in \mathcal{L}, k \in \mathcal{K}, \tag{5a}$$
$$\sum_{t\in\mathcal{T}}\sum_{k\in\mathcal{K}} B^k_{it}(\boldsymbol{\xi}) \ge \omega \sum_{r\in\mathcal{R}} y^r_i, \quad \forall i \in \mathcal{I}_\ell, \ell \in \mathcal{L}. \tag{5b}$$

Symmetry-Breaking Inequalities For multiple identical aircraft from the same service provider, their assignment decisions to OD pairs are interchangeable, which causes unnecessary larger decision space. To avoid redundant branch and bound time in exploring such equivalent solutions, this motivates the symmetry-breaking valid inequalities that cut off redundant alternative feasible solutions. For two identical aircraft $i, i' \in \mathcal{I}_\ell, i' \geq i+1, \ell \in \mathcal{L}$ with $m_i = m_{i'}, q_i^r = q_{i'}^r, c_{ij}^r = c_{i'j}^r, \alpha_{ij}^r = \alpha_{i'j}^r, \beta_{ij}^r = \beta_{i'j}^r$, we add the following inequalities to FairUAM (1):

$$\sum_{r\in\mathcal{R}} 2^r y_i^r \geq \sum_{r\in\mathcal{R}} 2^r y_{i'}^r \tag{6}$$

For two identical aircraft $i, i' \in \mathcal{I}_\ell, i' \geq i+1, \ell \in \mathcal{L}$ assigned to the same OD pairs, we can have the symmetry-breaking inequalities for different routes as follows:

$$\text{if } y_i^r = y_{i'}^r, \text{ then } \sum_{j\in\mathcal{J}_r} 2^j \delta_{ij}^r \geq \sum_{j\in\mathcal{J}_r} 2^j \delta_{i'j}^r, \quad \forall r \in \mathcal{R},$$

which can be equivalently represented as

$$\sum_{j\in\mathcal{J}_r} 2^j \delta_{i'j}^r - \sum_{j\in\mathcal{J}_r} 2^j \delta_{ij}^r \leq |y_i^r - y_{i'}^r| \mathcal{M}_r, \quad \forall r \in \mathcal{R}.$$

Above, for each route $r \in \mathcal{R}$, we let $\mathcal{M}_r$ be the upper bound of the left-hand side, where the value $2^{|\mathcal{J}_r|}$ suffices. To linearize the absolute operator, we can replace it with a variable $\bar{y}_{ii'}^r = |y_i^r - y_{i'}^r|$. In this case, we have $\bar{y}_{ii'}^r \geq |y_i^r - y_{i'}^r|$ and $\bar{y}_{ii'}^r \leq |y_i^r - y_{i'}^r|$. We can replace $\bar{y}_{ii'}^r \geq |y_i^r - y_{i'}^r|$ by introducing additional constraints:

$$\bar{y}_{ii'}^r \geq y_i^r - y_{i'}^r, \bar{y}_{ii'}^r \geq y_{i'}^r - y_i^r, \bar{y}_{ii'}^r \geq 0, \quad \forall r \in \mathcal{R}. \tag{7a}$$

Then, we introduce an additional binary variable $u_{ii'}^r \in \{0,1\}$ to linearize $\bar{y}_{ii'}^r \leq |y_i^r - y_{i'}^r|$ with constraints:

$$\bar{y}_{ii'}^r \leq y_i^r - y_{i'}^r + 2u_{ii'}^r, \bar{y}_{ii'}^r \leq y_{i'}^r - y_i^r + 2(1 - u_{ii'}^r), \quad \forall r \in \mathcal{R}. \tag{7b}$$

Since both variables $y_i^r, y_{i'}^r$ are binary, we can further relax the binary variable $u_{ii'}^r$ to be continuous by adding the following constraints:

$$1 - u_{ii'}^r \geq y_i^r - y_{i'}^r, u_{ii'}^r \geq y_{i'}^r - y_i^r, u_{ii'}^r \in [0,1], \quad \forall r \in \mathcal{R}. \tag{7c}$$

Finally, we arrive at the following symmetry-breaking inequalities for route assignment. For any two identical aircraft $i, i' \in \mathcal{I}_\ell, i' \geq i+1, \ell \in \mathcal{L}$ such that $m_i = m_{i'}, q_i^r = q_{i'}^r, c_{ij}^r = c_{i'j}^r, \alpha_{ij}^r = \alpha_{i'j}^r, \beta_{ij}^r = \beta_{i'j}^r$, we have

$$\sum_{j\in\mathcal{J}_r} 2^j \delta_{i'j}^r - \sum_{j\in\mathcal{J}_r} 2^j \delta_{ij}^r \leq \bar{y}_{ii'}^r \mathcal{M}_r, \quad \forall r \in \mathcal{R}, \tag{8}$$

(7a) – (7c).

Data Pre-processing Inequalities Since the auxiliary variables $\boldsymbol{B}(\boldsymbol{\xi})$ relate the aircraft location with bottleneck visiting time, their values can be one only between the earliest arrival time and latest arrival time for each bottleneck point. That is, we have the following pre-processing inequalities

$$B_{it}^k(\boldsymbol{\xi}) = 0, \quad \forall t < \underline{\tau}_{ik}, t \in \mathcal{T}, i \in \mathcal{I}_\ell, \ell \in \mathcal{L}, k \in \mathcal{K}, \tag{9a}$$

$$B_{it}^k(\boldsymbol{\xi}) = 0, \quad \forall t > \bar{L} + \bar{\tau}_{ik}, t \in \mathcal{T}, \forall i \in \mathcal{I}_\ell, \ell \in \mathcal{L}, k \in \mathcal{K}, \tag{9b}$$

where $\underline{\tau}_{ik} := \min_{r \in \mathcal{R}, j \in \mathcal{J}_r} \tau_{ijk}^r$ denotes the minimum time needed for aircraft $i \in I_\ell$ from service provider $\ell \in \mathcal{L}$ to arrive at bottleneck point $k \in \mathcal{K}$ in the free-condition airspace while $\bar{\tau}_{ik} := \max_{r \in \mathcal{R}, j \in \mathcal{J}_r} \tau_{ijk}^r$ denotes the maximum arrival time allowed. That is, before the minimum time needed for aircraft $i \in I_\ell$ from service provider $\ell \in \mathcal{L}$ to arrive at bottleneck point $k \in \mathcal{K}$ in the free-condition airspace, we must have $B_{it}^k(\boldsymbol{\xi}) = 0$; and after the maximum arrival time allowed at bottleneck point $k \in \mathcal{K}$, we also have $B_{it}^k(\boldsymbol{\xi}) = 0$.

Bottleneck Point Arrival Time Inequalities Similarly, the bottleneck arrival time for an aircraft is also related to its assignment. By definition, if $\sum_{r \in \mathcal{R}} y_i^r = 1$, we must have $\underline{\tau}_{ik} \leq \sum_{t \in \mathcal{T}} t B_{it}^k(\boldsymbol{\xi}) \leq \left(\bar{L} + \bar{\tau}_{ik}\right)$. Therefore, we can also bound the bottleneck arrival time based on OD pair assignment

$$\sum_{t \in \mathcal{T}} t B_{it}^k(\boldsymbol{\xi}) \leq \left(\bar{L} + \bar{\tau}_{ik}\right) \sum_{r \in \mathcal{R}} y_i^r, \quad \forall i \in \mathcal{I}_\ell, \ell \in \mathcal{L}, k \in \mathcal{K}, \tag{10a}$$

$$\sum_{t \in \mathcal{T}} t B_{it}^k(\boldsymbol{\xi}) \geq \underline{\tau}_{ik} \sum_{r \in \mathcal{R}} y_i^r, \quad \forall i \in \mathcal{I}_\ell, \ell \in \mathcal{L}, k \in \mathcal{K}. \tag{10b}$$

Total Arrival Time Inequalities We can upper bound the total arrival time of all aircraft at a specific bottleneck point based on its airspace capacity, which is

$$|\mathcal{T}| \max_{t \in \mathcal{T}} C_t^k \geq \sum_{\ell \in \mathcal{L}} \sum_{i \in \mathcal{N}_\ell} \sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{J}_r} \left(\tau_{ijk}^r \delta_{ij}^r + g_{ij}^r(\boldsymbol{\xi}) + \sum_{\substack{id(k) \geq id(k') \\ k' \in \mathcal{K}}} a_{ijk'}^r(\boldsymbol{\xi}) \right), \quad \forall k \in \mathcal{K}. \tag{11}$$

Accumulated Delay Inequalities When OD pair and route assignments (i.e., the first-stage decision variables) are given, an aircraft travels according to the predetermined bottleneck sequence. In other words, the bottleneck sequence depends on the first-stage assignment. Figure 2 is a demonstration, where we consider one OD pair with two alternative routes. In the first route, an aircraft visits bottleneck point $k_1 \in \mathcal{K}$ before bottleneck point $k_2 \in \mathcal{K}$, while it visits bottleneck point k_2 before bottleneck point k_1 in the second route.

In this case, there is no clear delay arrival time relationship between these two bottleneck points. Instead, we derive the relationship between the accumulated delays before visiting two bottleneck points. According to constraints (2g), the accumulated delay before visiting bottleneck point $k \in \mathcal{K}$

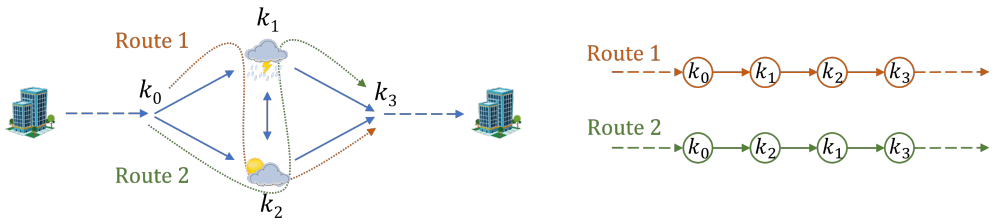

Figure 2 An illustration of one OD pair and two routes with different bottleneck visiting sequences

can be represented by the difference between the bottleneck actual arrival time $\sum_{t\in\mathcal{T}} tB^k_{it}(\boldsymbol{\xi})$ and the travel time needed in the free-condition airspace $\tau^r_{ijk}\delta^r_{ij}$ for any aircraft $i\in\mathcal{I}_\ell$ from service provider $\ell\in\mathcal{L}$ on route $j\in\mathcal{J}_r$ and OD pair $r\in\mathcal{R}$. Therefore, for any given route assignment with $\delta^r_{ij}=1$, when $id_j(k_1)<id_j(k_2)$ on route $j\in\mathcal{J}_r$ and OD pair $r\in\mathcal{R}$, we must have $\sum_{t\in\mathcal{T}} tB^{k_1}_{it}(\boldsymbol{\xi})-\tau^r_{ijk_1}\le \sum_{t\in\mathcal{T}} tB^{k_2}_{i,t}(\boldsymbol{\xi})-\tau^r_{ijk_2}$; on the other hand, we also note that $\sum_{t\in\mathcal{T}} tB^{k_1}_{it}(\boldsymbol{\xi})\le\sum_{t\in\mathcal{T}} tB^{k_2}_{i,t}(\boldsymbol{\xi})+L$ for any mixed-integer feasible solution. Thus, we obtain the following valid inequalities:

$$\sum_{t\in\mathcal{T}} tB^{k_1}_{it}(\boldsymbol{\xi})-\tau^r_{ijk_1}\delta^r_{ij}\le\sum_{t\in\mathcal{T}} tB^{k_2}_{i,t}(\boldsymbol{\xi})-\tau^r_{ijk_2}\delta^r_{ij}+(1-\delta^r_{ij})\bar{L},\quad \forall i\in\mathcal{I}_\ell,\ell\in\mathcal{L}. \tag{12}$$

We remark that inequalities (12) are very useful when we are given and want to evaluate a first-stage decision. Meanwhile, these inequalities can also strengthen the model formulation.

The effectiveness of inequalities (5) - (12) is demonstrated in Section 5.1 which further accelerates the algorithm proposed in the next section.

4.2 A Decomposition Algorithm

In this subsection, we present an exact decomposition algorithm to solve the FairUAM (1) together with valid inequalities. Our decomposition algorithm follows from Benders decomposition (Benders 1962). Note that in a conventional Benders decomposition, the original problem is partitioned into a master problem and a group of linear subproblems, which are much easier to solve for a given master solution. Following the same spirit, we partition FairUAM (1) into a master problem associated with OD pair and route assignment and a subproblem for ground and airborne delay assignment by relaxing the variables $\boldsymbol{B}(\boldsymbol{\xi})$ to be continuous. At each iteration, the upper bound can be obtained by plugging in the feasible delay assignment subproblem to the original formulation (1) and the lower bound can be computed by solving the master problem. It is known that the Benders cut is insufficient to recover the optimal value of FairUAM (1). To improve it, we adopt the L-shaped cut (Laporte and Louveaux 1993) which ensures the decomposition algorithm generates the exact optimal solution when terminated. Note that the L-shaped cut is generated by evaluating the subproblem with a given feasible first-stage solution.

Specifically, we let $\mathcal{X} = \{(\boldsymbol{\delta}, \boldsymbol{y}) : (1\text{b}) - (1\text{d})\}$ denote the feasible region of first-stage decision variables. By introducing an additional variable θ for the overall operation cost, we can formulate the *OD pair/route assignment master problem (MP)* as follows:

$$z_{mp} = \min_{\theta, \boldsymbol{\delta}, \boldsymbol{y}} \left\{ v + \varepsilon^{-1} \mathbb{E}_{\mathbb{P}}[h(\widetilde{\boldsymbol{\xi}})] : h(\widetilde{\boldsymbol{\xi}}) \geq \theta(\boldsymbol{\xi}) - v, (\boldsymbol{y}, \boldsymbol{\delta}) \in \mathcal{X}, (3\text{c}), (6) - (8) \right\} \tag{13}$$

For any fixed $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{\delta}}) \in \mathcal{X}$ and a realization of random parameter $\boldsymbol{\xi}$, the *delay assignment subproblem (SP)* is

$$\begin{aligned}
z_{sp}(\boldsymbol{\xi}) = \min_{\boldsymbol{g}(\boldsymbol{\xi}), \boldsymbol{a}(\boldsymbol{\xi}), \boldsymbol{s}(\boldsymbol{\xi}), \boldsymbol{B}(\boldsymbol{\xi})} \quad & Q(\bar{\boldsymbol{y}}, \bar{\boldsymbol{\delta}}, \boldsymbol{\xi}, \boldsymbol{g}(\boldsymbol{\xi}), \boldsymbol{a}(\boldsymbol{\xi}), \boldsymbol{s}(\boldsymbol{\xi}), \boldsymbol{B}(\boldsymbol{\xi})) && \text{(14a)} \\
\text{s.t.} \quad & B_{it}^k(\boldsymbol{\xi}) \in [0,1], \quad \forall i \in \mathcal{I}_\ell, \ell \in \mathcal{L}, k \in \mathcal{K}, t \in \mathcal{T}, && \text{(14b)} \\
& g_{ij}^r(\boldsymbol{\xi}), a_{ijk}^r(\boldsymbol{\xi}), s_\ell^r(\boldsymbol{\xi}) \geq 0, \quad \forall j \in \mathcal{J}_r, r \in \mathcal{R}, i \in \mathcal{I}_\ell, \ell \in \mathcal{L}, k \in \mathcal{K}, && \text{(14c)} \\
& (2\text{b}) - (2\text{i}), (4\text{b}) - (4\text{c}).
\end{aligned}$$

Let $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_8)$ be the dual variables associated with constraints (4c), (2b)-(2h) respectively. The *dual subproblem (DSP)* can be presented as follows:

$$\begin{aligned}
\max_{\boldsymbol{\pi}} \quad & \sum_{\ell \in \mathcal{L}} \left[\pi_{2\ell} \sum_{r \in \mathcal{R}} \sum_{i \in \mathcal{I}_\ell} \left(q_i^r \bar{y}_i^r + \sum_{j \in \mathcal{J}_r} c_{ij}^r \bar{\delta}_{ij}^r \right) + \pi_{3\ell} \left(D_\ell^r - \sum_{i \in \mathcal{I}_\ell} m_i \bar{y}_i^r \right) - \sum_{r \in \mathcal{R}} \sum_{i \in \mathcal{I}_\ell} \sum_{j \in \mathcal{J}_r} \bar{\delta}_{ij}^r \left(\pi_{4rij} \bar{L} + \pi_{5rij} \bar{R}_{ij}^r \right) \right. \\
& \left. + \sum_{i \in \mathcal{I}_\ell} \sum_{k \in \mathcal{K}} \left(-\pi_{6ik} + \pi_{7ik} \sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{J}_r} \tau_{ijk}^r \bar{\delta}_{ij}^r \right) \right] - \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \pi_{8kt} C_t^k, && \text{(15a)} \\
\text{s.t.} \quad & \sum_{\ell \in \mathcal{L}} \pi_{1\ell} |\mathcal{I}_\ell| \leq 1 - \lambda, && \text{(15b)} \\
& \pi_{2\ell} - \pi_{1\ell} \leq \frac{\lambda}{|I_\ell|}, \quad \forall \ell \in \mathcal{L}, && \text{(15c)} \\
& \pi_{3r\ell} - p_\ell^r \pi_{2\ell} \leq 0, \quad \forall r \in \mathcal{R}, \ell \in \mathcal{L}, && \text{(15d)} \\
& -\alpha_{ij}^r \pi_{2\ell} - \pi_{4rij} - \pi_{7ik} \leq 0, \quad \forall j \in \mathcal{J}_r, r \in \mathcal{R}, i \in \mathcal{I}_\ell, \ell \in \mathcal{L}, && \text{(15e)} \\
& -\beta_{ij}^r \pi_{2\ell} - \pi_{4rij} - \pi_{5ijr} - \pi_{7ik} \leq 0, \quad \forall j \in \mathcal{J}_r, r \in \mathcal{R}, i \in \mathcal{I}_\ell, \ell \in \mathcal{L}, k \in \mathcal{K} && \text{(15f)} \\
& -\pi_{6ik} + t\pi_{yik} - \pi_{8kt} \leq 0, \quad \forall i \in \mathcal{I}_\ell, \ell \in \mathcal{L}, k \in \mathcal{K}, t \in \mathcal{T}, && \text{(15g)} \\
& \pi_{1\ell}, \pi_{3r\ell}, \pi_{4rij}, \pi_{5ijr}, \pi_{6ik}, \pi_{8kt} \geq 0, \quad \forall j \in \mathcal{J}_r, r \in \mathcal{R}, i \in \mathcal{I}_\ell, \ell \in \mathcal{L}, k \in \mathcal{K}. && \text{(15h)}
\end{aligned}$$

If formulation (14) is feasible, we can obtain the optimal solution $\bar{\pi}$ of DSP (15) and the corresponding objective value $\bar{z}_{sp}$. Following the standard Benders decomposition procedure (Rahmaniani et al. 2017), the Benders cut for MP (13) can be generated as follows:

$$\begin{aligned}
\theta(\boldsymbol{\xi}) \geq & \bar{z}_{sp}(\boldsymbol{\xi}) + \sum_{\ell \in \mathcal{L}} \left[\bar{\pi}_{2\ell} \sum_{r \in \mathcal{R}} \sum_{i \in \mathcal{I}_\ell} \left(q_i^r (y_i^r - \bar{y}_i^r) + \sum_{j \in \mathcal{J}_r} c_{ij}^r (\delta_{ij}^r - \bar{\delta}_{ij}^r) \right) - \bar{\pi}_{3\ell} \sum_{i \in \mathcal{I}_\ell} m_i (y_i^r - \bar{y}_i^r) \right. \\
& \left. - \sum_{r \in \mathcal{R}} \sum_{i \in \mathcal{I}_\ell} \sum_{j \in \mathcal{J}_r} (\delta_{ij}^r - \bar{\delta}_{ij}^r) \left(\pi_{4rij} L + \pi_{5rij} R_{ij}^r \right) + \sum_{i \in \mathcal{I}_\ell} \sum_{k \in \mathcal{K}} \left(\pi_{7ik} \sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{J}_r} \tau_{ijk}^r (\delta_{ij}^r - \bar{\delta}_{ij}^r) \right) \right].
\end{aligned} \tag{16}$$

If formulation (14) or second-stage problem (2) is infeasible, we cut off the corresponding solution from MP (13) by adding the no-good cut (Codato and Fischetti 2006, Ahmed 2013) for given $\bar{\boldsymbol{\delta}}$ (we do not need to include $\bar{\boldsymbol{y}}$ according to constraints (1c)) as well as adding the feasibility cut

$$\sum_{r\in\mathcal{R}}\sum_{\ell\in\mathcal{L}}\sum_{i\in\mathcal{I}_\ell}\sum_{j\in\mathcal{J}_r}\left(\delta_{ij}^r(1-\bar{\delta}_{ij}^r)+(1-\delta_{ij}^r)\bar{\delta}_{ij}^r\right)\geq 1, \tag{17a}$$

$$\begin{aligned}0\geq\sum_{\ell\in\mathcal{L}}\Bigg[&\bar{\pi}_{2\ell}\sum_{r\in\mathcal{R}}\sum_{i\in\mathcal{I}_\ell}\left(q_i^r y_i^r+\sum_{j\in\mathcal{J}_r}c_{ij}^r\delta_{ij}^r\right)+\bar{\pi}_{3\ell}\left(D_\ell^r-\sum_{i\in\mathcal{I}_\ell}m_i y_i^r\right)-\sum_{r\in\mathcal{R}}\sum_{i\in\mathcal{I}_\ell}\sum_{j\in\mathcal{J}_r}\delta_{ij}^r\left(\bar{\pi}_{4rij}\bar{L}+\bar{\pi}_{5rij}\bar{R}_{ij}^r\right)\\&+\sum_{i\in\mathcal{I}_\ell}\sum_{k\in\mathcal{K}}\left(-\bar{\pi}_{6ik}+\bar{\pi}_{7ik}\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{J}_r}\tau_{ijk}^r\delta_{ij}^r\right)\Bigg]-\sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}}\bar{\pi}_{8kt}C_t^k,\end{aligned} \tag{17b}$$

where $\bar{\boldsymbol{\pi}}$ is an extreme direction of the dual problem. Note that even though no-good cut (17a) is commonly used in binary MP, feasibility cut (17b) is at least equally good. Thus, during our implementation, we use feasibility cut for the warm start and no-good cut and feasibility cuts for the branch and cut procedure.

Suppose that the MP solution $(\bar{\boldsymbol{y}},\bar{\boldsymbol{\delta}})$ is also feasible to the second-stage problem (2). Then we can evaluate its corresponding objective value by plugging into the second-stage problem (2) with valid inequalities:

$$\begin{aligned}z_e(\boldsymbol{\xi})=\min_{\boldsymbol{g}(\boldsymbol{\xi}),\boldsymbol{a}(\boldsymbol{\xi}),\boldsymbol{s}(\boldsymbol{\xi}),\boldsymbol{B}(\boldsymbol{\xi})}\ &Q(\bar{\boldsymbol{y}},\bar{\boldsymbol{\delta}},\boldsymbol{\xi},\boldsymbol{g}(\boldsymbol{\xi}),\boldsymbol{a}(\boldsymbol{\xi}),\boldsymbol{s}(\boldsymbol{\xi}),\boldsymbol{B}(\boldsymbol{\xi}))\\&\text{s.t.}\quad (2\text{b})-(2\text{k}),(4\text{b})-(4\text{c}),(5\text{a})-(5\text{b}),(9),(10),(12).\end{aligned} \tag{18}$$

We can obtain the optimal objective value $\bar{z}_e(\boldsymbol{\xi})$ for formulation (18), then the L-shaped cut for MP (13) can be generated as follows:

$$\theta(\boldsymbol{\xi})\geq z_e(\boldsymbol{\xi})+(z_e(\boldsymbol{\xi})-lb(\boldsymbol{\xi}))\left(\sum_{r\in\mathcal{R}}\sum_{\ell\in\mathcal{L}}\sum_{i\in\mathcal{I}_\ell}\sum_{j\in\mathcal{J}_r}\left(\delta_{ij}^r(1-\bar{\delta}_{ij}^r)+(1-\delta_{ij}^r)\bar{\delta}_{ij}^r\right)\right) \tag{19}$$

where $lb(\boldsymbol{\xi})$ is a valid lower bound of $\theta(\boldsymbol{\xi})$.

The detailed implementation can be found in Algorithm 1.

We can further warm-start Algorithm 1 by relaxing the integrality of $\boldsymbol{y}$ and $\boldsymbol{\delta}$ in MP (13) to generate optimality and feasibility cuts at the root node. The effectiveness of Algorithm 1 is demonstrated in Section 5.1.

The proposed inequalities (5) - (12) in Section 4 accelerate each iteration in Algorithm 1, improving the lower bound much faster compared to the original formulation. The convergence of the decomposition algorithm also improves a lot if we use both.

Algorithm 1: Decomposition Algorithm

Initialization: Set feasible region $\mathcal{X} = \{(\boldsymbol{y}, \boldsymbol{\delta}) : (1b) - (1d)\}, LB = 0, UB = \infty, \epsilon > 0$;
while $UB - LB > \epsilon$ **do**
  Solve the OD pair/route assignment master problem (MP) (13), record the optimal solution $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{\delta}})$ and optimal value z^*_{sp};
  $LB = \max\{LB, z^*_{mp}\}$;
  for each realization of random parameter $\boldsymbol{\xi}$ **do**
    Solve the delay assignment subproblem (SP) (14);
    if formulation (14) is feasible **then**
      Record the optimal value $z^*_{sp}(\boldsymbol{\xi})$ and add Benders cut (16) to MP (13);
      if formulation (18) is feasible **then**
        Record the optimal value $z^*_e(\boldsymbol{\xi})$ and add L-shaped cut (19) to MP (13)
      else
        Add no-good cut (17a) and feasibility cut (17b) to MP (13) and go to line 3;
  $UB = \min\{UB, \min_{v \in \mathbb{R}} \{v + \varepsilon^{-1} \mathbb{E}[z^*_e(\boldsymbol{\xi}) - v]_+\}\}$

5 Numerical Study

In this section, we present a group of numerical results to demonstrate the strengths of different model formulations and show the effectiveness of our solution method using random instances as well as a real-world network in Seattle. We report the results on random instances and conduct sensitivity analysis in Section 5.1. We then apply our solution method to a real-world network in Seattle in Section 5.2 for more aircraft fleet management insights. In reality, the UAM service providers operating on-demand services usually file their flight plan a day or hours before the actual departure time. We are allowed to obtain the allocation plan for the whole planning horizon offline. In this case, A time limit of 3,600 seconds was set for solving each random instance and the real-world case study. All the instances were coded in Python 3.7.0 with calls to Gurobi 9.0.3 on a personal laptop with a 1.9 GHz Intel Core i7 processor and 16 GB memory.

5.1 Results on Random Instances

Experimental Design and Setup We consider a small network and a larger network with different numbers of nodes and test 12 random instances for the small network and 8 random instances for the larger network. The small network has 5 vertiports, $|\mathcal{K}| = 5$ bottleneck points and $|\mathcal{R}| = 4$ OD pairs, while the larger network has 10 vertiports, $|\mathcal{K}| = 8$ bottleneck points and $|\mathcal{R}| = 6$ OD pairs. Each OD pair has 3 candidate routes. We consider $|\mathcal{L}| = 3$ service providers with different numbers of aircraft and each bears an aircraft capacity of $\boldsymbol{m} = 4$. The time horizon is $|\mathcal{T}| = 50$ and the travel time between 2 nodes in the free-condition airspace is set to either 2 or 3 time units. We let the

weighted scalar $\lambda = 0.5$. The aircraft relocation cost $\boldsymbol{q}$ is randomly selected from {0,1} to ensure that at most one origin vertiport among all OD pairs incurs a relocation cost of 0 if relocation is not needed. The assignment cost $\boldsymbol{c}$ is set equal to the travel time in a free condition. The ground delay cost $\boldsymbol{\alpha}$ is set to 1 and the airborne delay cost $\boldsymbol{\beta}$ is set to 3 per unit of time. The penalty cost $\boldsymbol{s}$ is set to 20 per unsatisfied passenger. The total travel delay limit is set to $\bar{L} = 10$ and the airborne delay limit is set $\bar{\boldsymbol{R}} = \mathbf{2}$. The airspace capacities of each bottleneck point per time unit are randomly generated with equal probability from two sets of different levels of airspace capacity: $\{1, 2\}$ for low level, $\{2, 3\}$ for high level. The passenger demand of each OD pair during the whole time horizon is generated from three different discrete uniform distributions to represent different levels of passenger demand: $\mathcal{U}(2, 6)$ for low level, $\mathcal{U}(6, 10)$ for medium level, $\mathcal{U}(10, 20)$ for high level. In the small network, the numbers of aircraft for service providers are set to 3 for the low-level passenger demand instances, 5 for the medium-level passenger demand instances and 10 for the high-level passenger demand instances. In the larger network, the numbers of aircraft for service providers are 6 for the low-level passenger demand instances, 10 for the medium-level passenger demand instances, 15 for the high-level passenger demand instances. We let the risk parameter $\varepsilon = 0.1$ in CVaR to minimize the worst case outcomes. The number of scenarios is chosen to be 5, 10, and 20 for the following reasons: i) the capacity uncertainty is due to inaccurate weather forecasting. However, there is a limited number of weather conditions. Hence, the scenarios that we consider are representative and can be of small size; ii) As mentioned in the previous section, the CVaR has distributional robustness interpretation, which is known to achieve better out-of-sample performance guarantees. We also illustrate the stability of the CVaR in Figure 4; and iii) Our FairUAM model is extremely difficult to solve. As shown in Table 1, using ten scenarios already takes a much longer time to solve to optimality.

Numerical Results We summarize the instances and report the results in Table 1 where a-b in the "Instance" column means the size network a and random instance b; "MILP" represents FairUAM (1) and "MILP.VI" represents FairUAM (1) with valid inequalities (5) - (12). We see that the computational time increases as the instance size (e.g., network size, number of aircraft, etc.) increases. When the number of scenarios increases, both objective value and computational time increases. MILP is only able to solve the first 5 instances to optimality. MILP.VI improves MILP's performance and can solve 80% more instances (i.e., the first 9 instances) to optimality. However, no instance in the larger network can be solved to optimality by either MILP or MILP.VI. On the contrary, Algorithm 1 can improve MILP.VI and solve 14 instances to optimality, which is 56% more instances solvable compared to MILP.VI and 180% more instances solvable compared to MILP. This is probably because using decomposition and exploring formulation structures take advantage of solving smaller and easier subproblems. Besides, we have found that the most effective cuts in

Algorithm 1 are L-shape and feasibility cuts. We also notice that, with the same passenger demand and the same number of aircraft, smaller airspace capacity instances are often more difficult to solve since the more restrictions in the airspace are, the harder aircraft assignment can be. When the airspace capacity increases, the objective value decreases since larger airspace capacity allows more aircraft to stay in the same airspace at the same time, which leads to short delays and smaller unsatisfied passenger demand. With the same airspace capacity, higher passenger demand results in longer computational time, since either more aircraft should be assigned to OD pairs or some UAM service providers may have unsatisfied passenger demand. Higher passenger demand also leads to a higher objective value, which is intuitive since the total operation cost of UAM service providers increases.

Table 1 Random Instance Setting and Results of Different Formulations and Algorithm 1

Instance	Num. of Scenarios	Airspace Capacity	Passenger Demand	MILP			MILP.VI			Algorithm 1		
				Time(s)	Gap(%)	Obj.Val	Time(s)	Gap(%)	Obj.Val	Time(s)	Gap(%)	Obj.Val
S-1	5	low	low	23	0	133.8	12	0	133.8	6	0.0	133.8
S-2			medium	67	0	285.8	33	0	285.8	12	0.0	285.8
S-3		high	medium	37	0	184.3	20	0	184.3	9	0.0	184.3
S-4			high	85	0	252.3	49	0	252.3	15	0.0	252.3
S-5	10	low	low	869	0	145.3	150	0	145.3	67	0.0	145.3
S-6			medium	3600	0.5	306.3	535	0	304.7	181	0.0	304.7
S-7		high	medium	3600	2.4	232.3	364	0	226.7	115	0.0	226.7
S-8			high	3600	2.7	352.7	1801	0	343.3	337	0.0	343.3
S-9	20	low	low	3600	2.2	189.8	2736	0	185.6	869	0.0	185.6
S-10			medium	3600	5.4	379.3	3600	0.3	359.8	2497	0.0	358.8
S-11		high	medium	3600	5.3	285.2	3600	0.8	272.3	1684	0.0	270.0
S-12			high	3600	6.8	430.0	3600	1.0	406.7	3600	0.2	403.5
L-13	10	low	low	3600	13.4	292.2	3600	3.9	264.6	666	0.0	254.3
L-14			medium	3600	17.3	340.1	3600	5.4	299.1	1631	0.0	282.9
L-15		high	medium	3600	15.6	353.3	3600	5.2	313.3	1055	0.0	296.7
L-16			high	3600	25.1	714.5	3600	8.6	603.4	3600	0.4	549.9
L-17	20	low	low	3600	38.9	441.9	3600	11.3	324.4	3600	1.3	294.9
L-18			medium	3600	46.1	717.4	3600	14.9	489.1	3600	4.4	435.4
L-19		high	medium	3600	46.9	609.6	3600	15.3	396.6	3600	7.3	362.2
L-20			high	3600	57.3	904.7	3600	21.4	681.0	3600	10.2	596.0

To better demonstrate the effectiveness of valid inequalities, besides comparing computational time and optimality gap of MILP and MILP.VI in Table 1, we solved instance S-4 using Algorithm 1 with and without valid inequalities (5) - (12) respectively, which are illustrated in Figure 3. Note that, the results for other instances are similar and Figure 3 is for demonstration. The computational time of instance S-4 significantly decreases from 438 seconds to 15 seconds if we apply valid inequalities. The improvement mainly comes from the symmetry-breaking ones that prevent exploring unnecessary equivalent solutions.

To demonstrate the robustness of the solutions from our model, we consider an additional risk-neutral FairUAM (1) where $\varepsilon = 1$ and $\rho(Z) = \mathbb{E}_{\mathbb{P}}[Z]$ to compare with our risk-averse one. We first

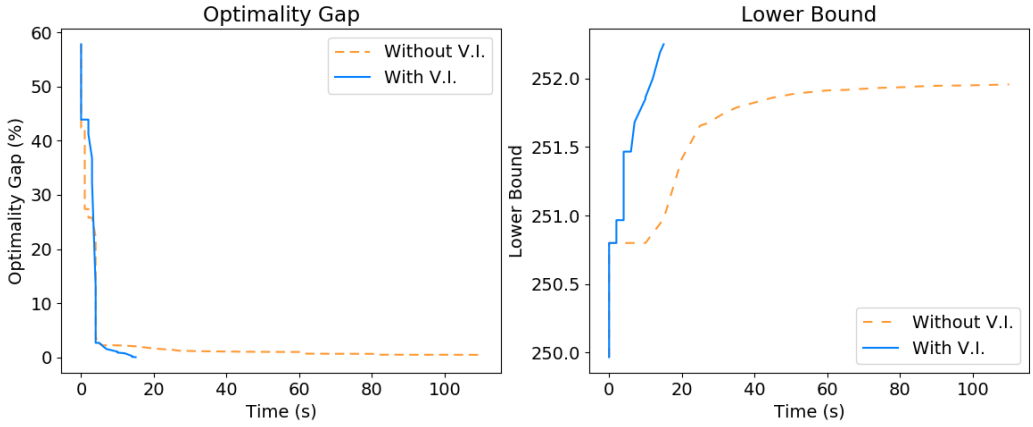

Figure 3 Comparison of Lower Bound and Optimality Gap with and without the Valid Inequalities

obtained the solutions by solving two formulations with instance S-8. Then, we generated 10 new scenarios following the same procedure mentioned before for evaluation. We assumed that weather conditions or passenger demand are changing rapidly over time. Thus, we applied a truncated Gaussian noise to generate airspace capacity and passenger demand per scenario to make sure that they are nonnegative. More specifically, we let the random perturbation of airspace capacity follow a Gaussian distribution $\mathcal{N}(0, 0.1)$ truncated to be non-positive and rounded to the nearest nonnegative integer. For the random perturbation applied to passenger demand, we let it follow a Gaussian distribution $\mathcal{N}(0, 0.2)$ truncated to a nonnegative number and rounded to the nearest integer. We repeated the process 15 times to generate asymptotic 95% confidence intervals. The results are illustrated in Figure 4. It shows that, when the demand increases or the airspace capacity decreases, the objective value increases and the risk-averse one is more robust. It is worth mentioning that, when the airspace capacity is small (e.g., a part of bottleneck points having 0 capacity), no aircraft can depart; otherwise, it will violate the airspace capacity. In this case, both risk-neutral and risk-averse models end up with a trivial solution.

We conducted a sensitivity analysis of weighted scalar λ on instance S-8 to demonstrate the trade-off between efficiency and fairness. We let $\lambda \in \{0, 0.25, 0.5, 0.75, 1\}$. For better comparison, we illustrate the change in percentage for total cost and unfairness, where for the total cost change, we let the base case be the lowest possible cost, and for the unfairness change, we let the base case be the smallest value of the largest average company cost. We evaluate unfairness as the absolute difference between the highest and lowest average company operation costs. The result is illustrated in Figure 5. We see that, when the weighted scalar $\lambda = 1$, the total operation cost is the smallest; however, the solution is extremely unfair among different UAM service providers. Some service providers suffer much higher average aircraft operation cost due to delays. As the weighted scalar increases, we can reduce the unfairness dramatically with a small increase in the total operation

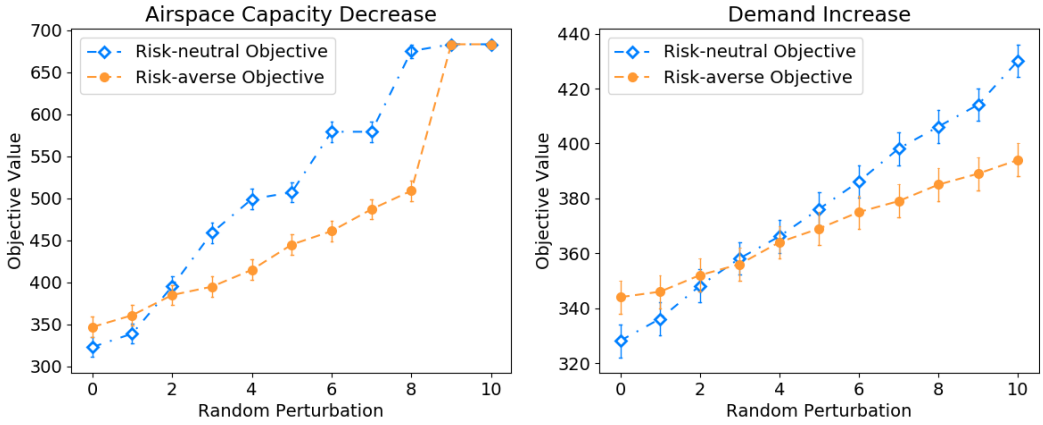

Figure 4 Evaluation of the Objective Value when Airspace Capacity Decreases or Demand Increases

cost. When the weighted scalar $\lambda = 0$, we make the fairest assignment for each service provider; however, as expected, the total operation cost is the highest. In summary, we recommend using an appropriate parameter λ which can help achieve a good balance between fairness and efficiency.

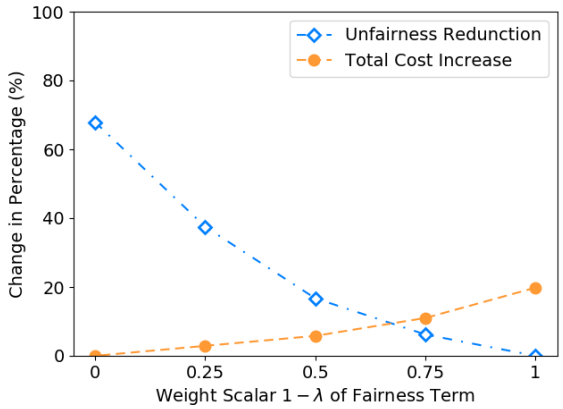

Figure 5 Trade-off between Fairness and Efficiency

To illustrate that a small number of scenarios are sufficient to generate effective decisions, we conducted a numerical study on four small networks with a time horizon $|\mathcal{T}| = 20$ to evaluate the out-of-sample performance of solutions obtained from 20 and 100 scenarios. In this numerical study, each network consists of $|\mathcal{L}| = 3$ service providers and $|\mathcal{R}| = 3$ OD pairs, and each service provider operates three routes. Networks 1 and 2 have $|\mathcal{K}| = 3$ bottleneck points, with each OD pair having 3 candidate routes. The airspace capacity follows a discrete uniform distribution of $\mathcal{U}(1, 2)$, and the passenger demand follows a discrete uniform distribution of $\mathcal{U}(2, 6)$. Networks 3 and 4 have $|\mathcal{K}| = 5$ bottleneck points, with each OD pair having 5 candidate routes. The airspace capacity follows a discrete uniform distribution of $\mathcal{U}(3, 4)$, and the passenger demand follows a

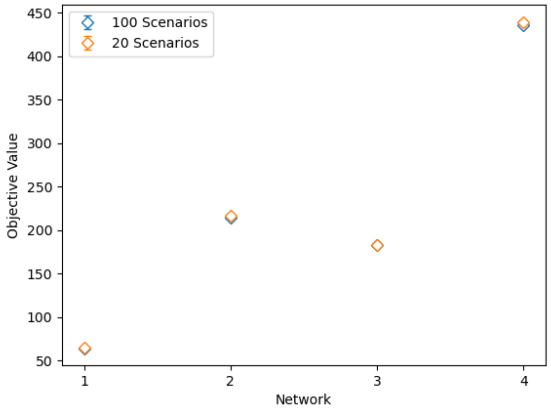

Figure 6 Out-of-Sample Performances of the Decisions obtained from 20 and 100 Scenarios

Table 2 A Summary of Out-of-Sample Performance Results from 20 and 100 Scenarios

Network	Num. of Scenarios	Time(s)	Gap(%)	Obj. Val	95% Confidence Interval
1	20	31.2	0.0	58.0	[63.3, 65.2]
1	100	1590.7	0.0	60.3	[63.1, 64.9]
2	20	14.2	0.0	198.3	[215.2, 217.7]
2	100	32.7	0.0	212.3	[213.6, 216.1]
3	20	9.9	0.0	169.5	[181.1, 183.8]
3	100	57.0	0.0	183.4	[181.4, 184.7]
4	20	2.1	0.0	427.8	[437.6, 441.1]
4	100	10.7	0.0	440.7	[433.6, 437.5]

discrete uniform distribution of $\mathcal{U}(10, 20)$. First, for each network, we obtained first-stage decisions by solving the problem with 20 and 100 scenarios. Next, we generated 100 new samples following the distribution of airspace capacity and passenger demand of the corresponding network to evaluate these decisions. This procedure was repeated 100 times for each network and scenario size to compute asymptotic 95% confidence intervals, as depicted in Figure 6 and Table 2. It is seen that the decisions obtained from 20 and 100 scenarios are of similar out-of-sample performances. Moreover, solving the problem for 100 scenarios requires a significantly longer computational time. This confirms that a small number of scenarios are adequate to solve the proposed model. This also validates the efficacy of the risk measure- CVaR.

To further justify the selection of CVaR as the risk measure, we compare it with Value-at-Risk (VaR) (see, e.g., Rockafellar et al. 2000), another popular risk measure. Specifically, we solved network 3 using 20 scenarios to obtain first-stage decisions for both risk measures with $\epsilon = 0.1$. Then, we generated 100 samples 20 times and applied random perturbations to the airspace capacity and passenger demand following the same procedure as previously described to generate 95% confidence intervals for our analysis. The results are illustrated in Figure 7. The comparison reveals that

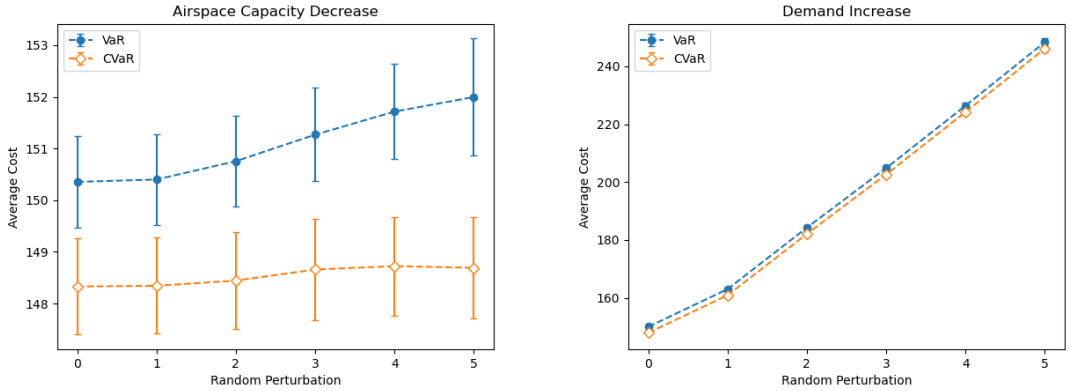

Figure 7 Comparison of CVaR and VaR models

VaR and CVaR demonstrate similar robustness, while CVaR results in a lower average out-of-sample cost. We also compared the computational time of CVaR and VaR as risk measures by solving networks 1-4 using 20 scenarios. The results are presented in Table 3. Notably, due to the nonconvexity of VaR, it often requires a significantly longer time to solve.

Table 3 A Comparison of CVaR and VaR

Network	Num. of Scenarios	Risk Measure	Time(s)	Gap(%)	Obj.Val
1	5	CVaR	7.4	0.0	54.7
		VaR	8.9	0.0	54.7
	10	CVaR	9.4	0.0	61.2
		VaR	18.8	0.0	57.8
	20	CVaR	31.2	0.0	58.0
		VaR	327.3	0.0	51.8
2	5	CVaR	0.8	0.0	202.3
		VaR	4.7	0.0	202.3
	10	CVaR	8.7	0.0	209.3
		VaR	7.8	0.0	202.7
	20	CVaR	14.1	0.0	198.3
		VaR	504.4	0.0	195.2
3	5	CVaR	7.7	0.0	160.0
		VaR	30.2	0.0	160.0
	10	CVaR	14.5	0.0	165.2
		VaR	48.9	0.0	159.0
	20	CVaR	10	0.0	169.5
		VaR	157	0.0	156.2
4	5	CVaR	0.8	0.0	423.0
		VaR	5.2	0.0	423.0
	10	CVaR	1.1	0.0	423.0
		VaR	14.1	0.0	422.7
	20	CVaR	2.1	0.0	427.8
		VaR	22.4	0.0	416.3

5.2 Case Study on a Real-World Network

Experimental Design and Setup We consider 13 airports within 50 miles from Seattle-Tacoma International Airport (KSEA), where their airport codes are denoted as the location of vertiports. The connectivity of vertiports is shown in Figure 8. The airspace above each vertiport is considered an intersection point of different routes. For example, two possible routes for OD pair (KSEA, KSHN) are: (Route 1) KSEA $\rightarrow$ 2S1 $\rightarrow$ KSHN, (Route 2) KSEA $\rightarrow$ KTIW $\rightarrow$ KSHN. Aircraft assigned to these routes will enter the airspace above KSEA, 2S1, KTIW, and KSHN, considered the bottleneck points on these routes. Note that we use the same name to represent the airspace. We let the time horizon $|\mathcal{T}|$ be 60 and a single unit of time be equal to 2 minutes, that is, the planning horizon is 2 hours in total. Travel time needed between 2 nodes in the free-condition airspace is set equal to the Euclidean distance divided by the travel speed 4 miles per time unit and rounded up to the nearest integer, i.e., travel time between KBFI and KSEA (the Euclidean distance is 5.6 miles) is 2 units of time. We consider $|\mathcal{L}| = 3$ service providers, each with a different number of aircraft and having different aircraft with capacity varying from $\boldsymbol{m} \in \{3, 4, 5\}$. Each service provider corresponds to 10 OD pairs, and all 25 OD pairs are covered by at least one service provider. If an OD pair is operated by more than one service provider, the demand is assigned evenly to each service provider. The aircraft relocation cost $\boldsymbol{q}$ is selected from $\{0, 1, 2\}$ with equal probability. The assignment cost $\boldsymbol{c}$ is set to be equal to the travel time in a free condition multiplied by the unit cost 0.5 per time unit. The ground delay cost $\boldsymbol{\alpha}$ is set to 1, and airborne delay cost $\boldsymbol{\beta}$ is set to 3 per unit of time. The penalty cost $\boldsymbol{s}$ is 20 per unsatisfied passenger. We make this assumption because one commonly used air-ground delay cost ratio in the literature is 3 (Mukherjee and Hansen 2009). We set the weighted parameter $\lambda = 0.5$ to balance the fairness and the efficiency, and risk parameter $\varepsilon = 0.1$ to hedge against the risk. We also use results from the model without fairness ($\lambda = 0, \varepsilon = 0.1$) and the risk-neutral model ($\lambda = 0.5, \varepsilon = 1$) for comparison.

Data Input The airspace capacity of each bottleneck point was generated based on the Doppler radar map on Nov 14th, 2021, from 6 PM to 8 PM. The full capacity is set to 3 aircraft per time unit when the color is gray on the map for corresponding vertiport locations. The reduced capacity is set to 2 or 1 aircraft per time unit when the color becomes blue or green, respectively. To model the uncertain parameters, we shifted the timeline by 5 units for each scenario and applied a truncated Gaussian distribution $\mathcal{N}(0, 0.5)$ rounded to the nearest integer to generate the airspace capacity of each bottleneck point.

The passenger demand was generated based on population and airport operations data. We first collected the total population data in the corresponding area having the same zipcode for each vertiport then multiplied them by 0.02%, assuming that 0.02% of the population plan to travel

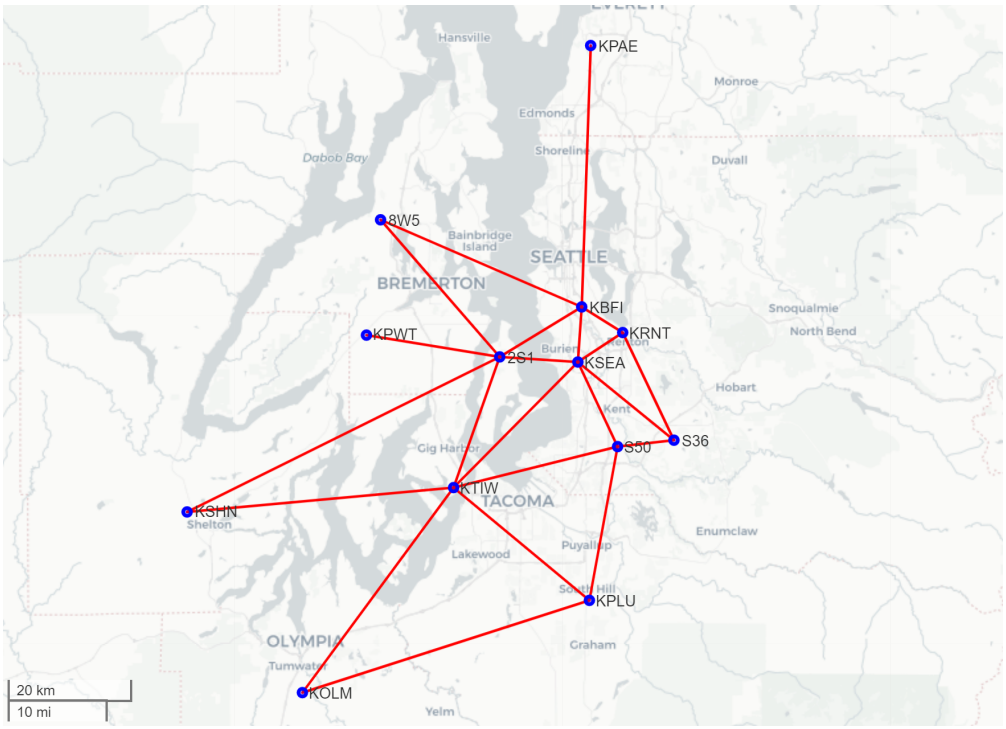

Figure 8 Map of 13 airports within 50 miles from KSEA

during the planning horizon. We also collected the average transient general aviation operation data for each vertiport, and multiplied them by 0.02 to represent the travel needs of people not living in the corresponding area. We let the sum of these two numbers for each vertiport to represent the total passenger demand of all the OD pairs that depart from the same vertiports. For a given origin vertiport, its destination vertiports are the nearest two vertiports at least 30 miles away and within 50 miles. (An origin vertiport may have only one destination vertiport, i.e., KTIW has only one destination.) We then assigned the total passenger demand to OD pairs evenly and rounded them up to the nearest integers. For example, KSEA has two destination vertiports, KPAE (31.6 miles) and KSHN (41.9 miles). Each OD pair has a passenger demand of 13 since the total passenger demand at KSEA is 25. To model the future uncertainty, we applied a truncated Gaussian distribution $\mathcal{N}(0, 2)$ rounded to the nearest nonnegative integer to the passenger demand of each OD pair. In our numerical study, we generated 10 scenarios for this problem.

Numerical Results We ran Algorithm 1 to solve all the cases in this subsection and display the results in Table 4, where "Max" denotes the largest average company cost, "Ave.G" denotes the average company ground delay among different scenarios, "Ave.A" denotes the average company airborne delay, and "Ave.S" denotes the average unsatisfied passenger demand. "Num. of Aircraft" denotes the number of aircraft of each service provider, which corresponds to the planning horizon. The computational time of all the cases is less than an hour. It is seen that increasing the number

of aircraft can decrease the unsatisfied passenger demand and optimal value. And all three models yield the same unsatisfied passenger demand. This is probably because satisfying the passenger demand is prioritized in all models. Although the ratio of the cost of airborne delay and ground delay is not large, airborne delay tends to be close to zero and ground delay is always larger than zero. This is not surprising, since when the delay cannot be avoided, it is better to hold aircraft on the ground instead of letting them wait in the air. Compared to the model without fairness, the model with fairness can decrease the highest company operation cost by about 10% with a 2% increase in the optimal value. Notice that, in the fair and risk-averse model, due to fairness enforcement among different UAM service providers, those service providers with nonzero unsatisfied passenger demand are assigned to have a smaller ground delay. However, in the model without fairness, due to lack of the fairness term, a service provider with nonzero unsatisfied passenger demand may be assigned to a large ground delay, which leads to a larger company operation cost. Furthermore, the largest ground delay in the fair and risk-averse model is always higher than that from the model without fairness. Compared to the result from the risk-neutral model, the risk-averse one yield a small increase of 5% in the total operation cost and the largest company operation cost since it is optimizing the worst scenario performance instead of average performance over all the scenarios. To further demonstrate the robustness of solutions from the two models, following the same procedure in Section 5.1, we generated 5 new scenarios and applied a truncated Gaussian noise to passenger demand. The result is shown in Figure 9, which shows that, the total operation cost increases and the risk-averse one tends to be more robust when the passenger demand increases.

Table 4 Results of Case Study on A Real-World Network

Case	Num. of Aircraft	Fair and Risk Averse					Without Fairness					Risk Neutral				
		Total	Company Cost				Total	Company Cost				Total	Company Cost			
			Max	Ave.G	Ave.A	Ave.S		Max	Ave.G	Ave.A	Ave.S		Max	Ave.G	Ave.A	Ave.S
1	(10,10,10)	40.4	14.6	(16,4,5)	(0,0,0)	(5,6,4)	39.7	15.8	(4,18,2)	(0,0,0)	(5,6,4)	38.3	14.2	(13,4,4)	(0,0,0)	(5,6,4)
2	(15,15,15)	17.1	6.7	(15,25,17)	(0,0,0)	(4,0,0)	16.7	7.3	(27,13,15)	(0,0,0)	(4,0,0)	16.5	6.4	(15,22,16)	(0,0,0)	(4,0,0)
3	(18,15,12)	12.5	4.2	(24,23,8)	(0,0,0)	(0,0,2)	12.2	4.9	(10,25,17)	(0,0,0)	(0,0,2)	11.9	4.1	(21,20,8)	(0,0,0)	(0,0,2)

Managerial Insight FairUAM can provide the UAM traffic manager with an optimal UAM aircraft resource allocation plan and delay assignment without perfect information, achieving both company-level fairness and system-level efficiency. The framework accounts for the existence of uncertainties in weather information and demand fluctuation. Besides generating an optimal operation plan, our framework can also provide extra support for the UAM traffic manager to identify the busy bottleneck points, which are usually regarded as the system hotspots. With such information, UAM service providers are capable of better designing their routes in the aircraft path-planning

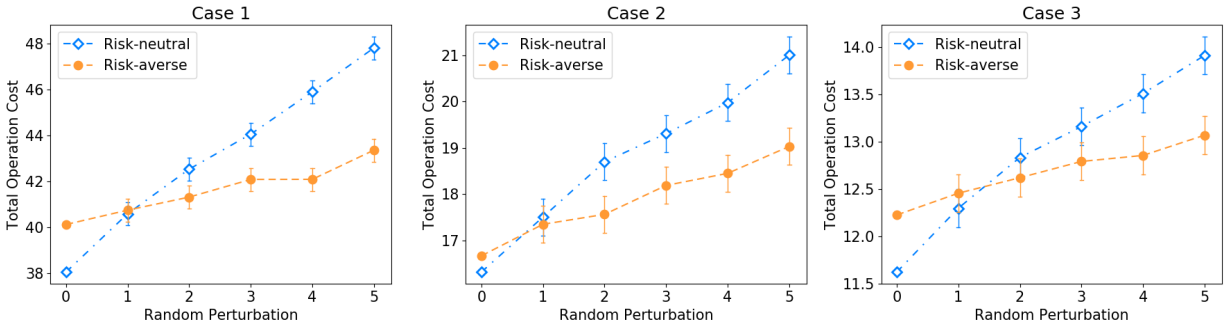

Figure 9 Evaluation of Total Operation Cost when Passenger Demand Increases

stage before filing the flight plan to the UAM traffic manager in the future. UAM service providers can consider more candidate routes or even take detours to avoid busy bottleneck points and possible future congestion. In addition, identification of the current system hotspots can also be valuable when the UAM network is still in the blueprint or can improve the existing UAM network. Instead of improving the capability of all the bottleneck points, which is quite expensive, focusing on the busiest bottleneck points is more effective in a mature UAM operating environment with high-density aircraft. In this case, we would like to evaluate the utilization rate of bottleneck points. The higher the utilization rate is, the busier the corresponding bottleneck point should be.

Suppose in time horizon $\mathcal{T}$, the time point that the last aircraft landing at its destination vertiport is

$$t_0 = \arg\min_{t'} \left\{ t' : \sum_{t>t', t\in\mathcal{T}} tB_{it}^k(\boldsymbol{\xi}) = 0, \forall k \in \mathcal{K} \right\}.$$

We consider the utilization rate of the bottleneck point $k \in \mathcal{K}$ between time 0 and t_0 as the ratio of the actual number of aircraft passing the bottleneck point divided by the total number of aircraft it can handle during this period, which can be represented as

$$UR_k = \frac{\sum_{t\in\mathcal{T}_0}\sum_{\ell\in\mathcal{L}}\sum_{i\in\mathcal{I}_\ell} B_{it}^k(\boldsymbol{\xi})}{\sum_{t\in\mathcal{T}_0} C_t^k}$$

where $\mathcal{T}_0 = \{0, \ldots, t_0\}$. The results are displayed in Table 5. We see that KSEA, KBFI, 2S1, KTIW are the busiest bottleneck points with the highest utilization rate. This may be because these bottleneck points are located around the center of the network, and many routes pass those bottleneck points more often compared to others. When we numerically increase the total airspace capacity $\sum_{t\in\mathcal{T}_0} C_t^k$ of those bottleneck points by 10% in Case 1, we find that the optimal value can be reduced by 5%. Note that the results of Cases 2 and 3 are similar. This implies that if the traffic manager would like to further reduce total congestion or operation cost, it is a good idea to focus on these four bottleneck points. Enlarging the capability of these four bottleneck points can benefit the whole network. Note that the capacity of the system hotspots can be increased

by introducing promising advanced aircraft separation assurance technologies and aircraft onboard automation (see, e.g., Kochenderfer et al. 2012, Brittain et al. 2020). We further illustrated how the total operation cost and the fairness change while increasing the total capacity of these four bottleneck points, as shown in Figure 10. The fairness score is defined as the ratio of the smallest average company cost to the largest average company cost such that 1 is completely fair and 0 is completely unfair. It is shown that the total operation cost decreases when the total capacity increases since the congestion level can be reduced. However, when the total capacity of these four bottleneck points is increased by over 15%, the objective value tends to stay at a similar value since the system hotspots transfer to other bottleneck points.

On the other hand, changing the number of bottleneck points on a route when designing the routes, i.e., taking a detour to avoid visiting busy airspace, can also decrease the delay. We demonstrate the impact on the total operation cost and the largest company average operation cost when some routes are redesigned to avoid busy bottleneck points (KSEA, KBFI, 2S1, KTIW) using Case 1, which is shown in Figure 10. We can see that the total operation cost can be reduced, while the fairness almost stays the same. This suggests that to achieve the best efficiency, it may not have to sacrifice fairness between UAM service providers. The overall efficiency can also be improved with better network and route design to obtain a win-win situation.

Table 5 Airspace Utilization Rate

Case	KSEA	KRNT	KBFI	2S1	S50	S36	KTIW	KPWT	KPLU	8W5	KPAE	KSHN	KOLM
1	0.33	0.14	0.25	0.31	0.17	0.08	0.25	0.11	0.19	0.17	0.11	0.31	0.17
2	0.50	0.20	0.40	0.40	0.25	0.10	0.50	0.15	0.30	0.30	0.15	0.35	0.25
3	0.55	0.25	0.50	0.60	0.30	0.15	0.45	0.15	0.35	0.30	0.20	0.20	0.25

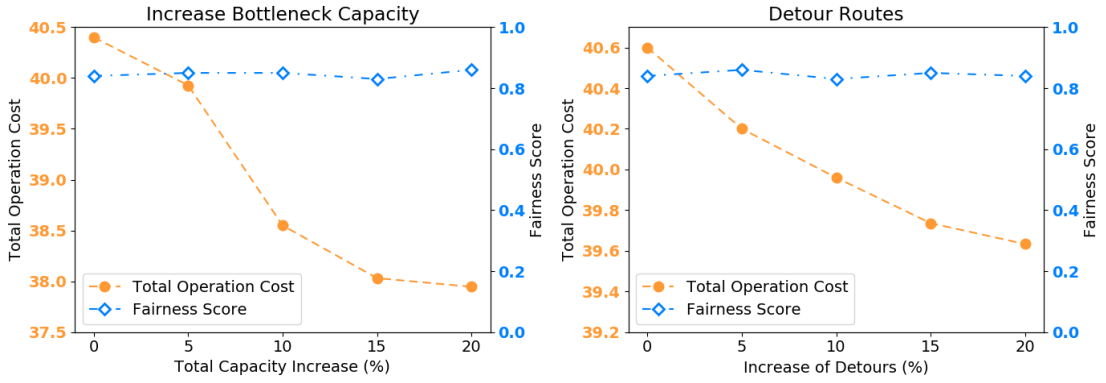

Figure 10 Impact on Total Operation Cost and Fairness while Increasing Total Capacity of Busy Bottleneck Points (KSEA, KBFI, 2S1, KTIW) or Taking Detours to Avoid Them

In our FairUAM framework, a better input (e.g., more accurate weather prediction) leads to a better operation plan and less computational effort. Following the same procedure in applying the random Gaussian perturbation to airspace capacity, we demonstrate the impact of the accuracy of airspace capacity on the total operation cost and the largest average company operation cost using Case 1 as shown in Figure 11. We notice that by increasing the fluctuation in airspace capacity, the total operation cost increases. This raises the requirement for a better weather forecast or even an online weather report for each bottleneck point. For the passenger aspect, changes in passenger demand result in an extra computational effort or sometimes unsatisfied demand. Therefore, instead of only focusing on launching the service or developing the technology, it is also important for the service providers to understand their passengers and incentivize the passengers for higher service satisfaction.

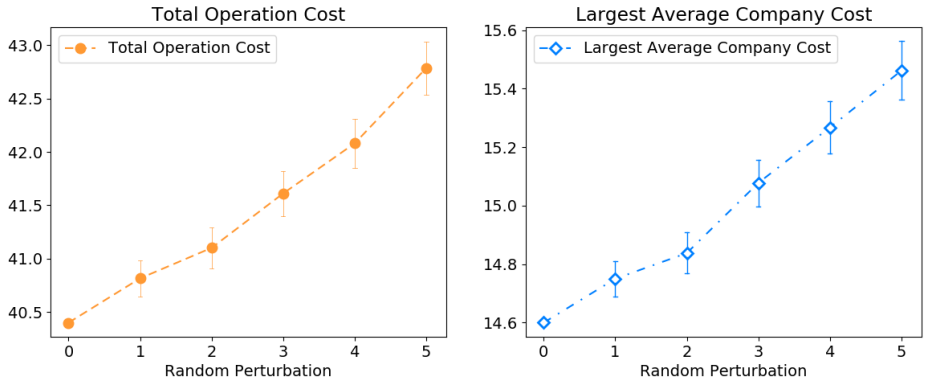

Figure 11 Evaluation of Total Operation Cost and Largest Average Company Cost when Airspace Capacity Changes

6 Conclusion

In this paper, we study the FairUAM, which is proven to be NP-hard in a deterministic setting and NP-hard with a given route and OD pair assignment. To simplify FairUAM, we derive monotonicity properties and relax the integrality of some decision variables. To further improve the MILP formulation, we propose valid inequalities by exploring the model structures. We develop a decomposition-based algorithm to solve FairUAM. Finally, we generate random instances to demonstrate the effectiveness of the proposed method to apply it to a real-world network in Seattle. Compared to the risk-neutral model and the model without fairness consideration, FairUAM is more robust when the passenger demand or weather forecasting is subject to error and can fairly assign the aircraft and delay at the company-level. Aircraft-level fairness can also be included in FairUAM with a simple modification to the fairness term. For the managerial aspect, FairUAM

not only generates the optimal operation plan but also detects the system hotspots by finding the busy bottleneck points. This information provides UAM service providers extra support for better route planning. It also provides some guidance in future UAM network design. This is consistent with the UAM concept of operations provided by the Federal Aviation Administration (FAA) and National Aeronautics and Space Administration (NASA) (Fontaine 2023) as

> "...The 'buffer' necessary to account for uncertainty as the operational tempo increases leads to the eventual need for tactical deconfliction and DCB (demand capacity balancing) capabilities to optimize efficiency..."

For our future work, we would like to include network design and rerouting (see, e.g., Yu et al. 2021b, Toth and Vigo 2002) into consideration and theoretically improve the model using stronger formulations. The FairUAM framework can also be generalized to other resource allocation problems under demand and capacity uncertainties. For example, we can extend our framework to robot task allocation with demand and capacity uncertainties in the warehouse. We can also generalize the framework to fair vehicle allocation in disaster evacuation to minimize travel time.

Acknowledgments

The first, second, and fourth authors were supported in part by the National Science Foundation grants 2246414 and 2246417. The third author was supported by the National Science Foundation grant 2047390. The fourth author was also supported by the Virginia Space Grant Consortium New Investigator Program.

References

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

Benders, J.F. 1962. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4 (1), 238-252.

Bennaceur, Mehdi, Rémi Delmas, Youssef Hamadi. 2022. Passenger-centric urban air mobility: Fairness trade-offs and operational efficiency. *Transportation Research Part C: Emerging Technologies*, 136 103519.

Bertsimas, Dimitris, Guglielmo Lulli, Amedeo Odoni. 2011. An integer optimization approach to large-scale air traffic flow management. *Operations research*, 59 (1), 211-227.

Bharadwaj, Suda, Steven Carr, Natasha Neogi, Hasan Poonawala, Alejandro Barberia Chueca, Ufuk Topcu. 2019. Traffic management for urban air mobility. *NASA Formal Methods Symposium*. Springer, 71-87.

Bharadwaj, Suda, Tichakorn Wongpiromsarn, Natasha Neogi, Joseph Muffoletto, Ufuk Topcu. 2021. Minimum-violation traffic management for urban air mobility. *NASA Formal Methods Symposium*. Springer, 37-52.

Brittain, Marc, Xuxi Yang, Peng Wei. 2020. A deep multi-agent reinforcement learning approach to autonomous separation assurance. *arXiv preprint arXiv:2003.08353*, .

Chen, Jun, Long Chen, Dengfeng Sun. 2017. Air traffic flow management under uncertainty using chance-constrained optimization. *Transportation Research Part B: Methodological*, 102 124-141.

Chin, Christopher, Karthik Gopalakrishnan, Maxim Egorov, Antony Evans, Hamsa Balakrishnan. 2021. Efficiency and fairness in unmanned air traffic flow management. *IEEE Transactions on Intelligent Transportation Systems*, 22 (9), 5939-5951.

Codato, Gianni, Matteo Fischetti. 2006. Combinatorial benders' cuts for mixed-integer linear programming. *Operations Research*, 54 (4), 756-766.

Connors, Mary M. 2020. Understanding risk in urban air mobility: Moving towards safe operating standards, .

Daskilewicz, Matthew, Brian German, Matthew Warren, Laurie A Garrow, Sreekar-Shashank Boddupalli, Thomas H Douthat. 2018. Progress in vertiport placement and estimating aircraft range requirements for evtol daily commuting. *2018 Aviation Technology, Integration, and Operations Conference*. 2884.

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

Faghih-Roohi, Shahrzad, Yew-Soon Ong, Sobhan Asian, Allan N Zhang. 2016. Dynamic conditional value-at-risk model for routing and scheduling of hazardous material transportation networks. *Annals of Operations Research*, 247 715-734.

Fontaine, Paul. 2023. Urban air mobility (uam) concept of operations v2.0. URL `https://www.faa.gov/sites/faa.gov/files/Urban%20Air%20Mobility%20%28UAM%29%20Concept%20of%20Operations%202.0_0.pdf`.

Ganji, Moein, David J Lovell, Michael O Ball, Alex Nguyen. 2009. Resource allocation in flow-constrained areas with stochastic termination times. *Transportation research record*, 2106 (1), 90-99.

Garrow, Laurie A, Brian J German, Caroline E Leonard. 2021. Urban air mobility: A comprehensive review and comparative analysis with autonomous and electric ground transportation for informing future research. *Transportation Research Part C: Emerging Technologies*, 132 103377.

Graham, Ronald L. 1966. Bounds for certain multiprocessing anomalies. *Bell system technical journal*, 45 (9), 1563-1581.

Gupta, Shubham, Dimitris J Bertsimas. 2011. Multistage air traffic flow management under capacity uncertainty: a robust and adaptive optimization approach, .

Hamdan, Sadeque, Ali Cheaitou, Oualid Jouini, Tobias Andersson Granberg, Zied Jemai, Imad Alsyouf, Maamar Bettayeb, Billy Josefsson. 2022. Central authority–controlled air traffic flow management: An optimization approach. *Transportation Science*, 56 (2), 299-321.

Hamdan, Sadeque, Ali Cheaitou, Oualid Jouini, Zied Jemai, Imad Alsyouf, Maamar Bettayeb. 2018. On fairness in the network air traffic flow management with rerouting. *2018 9th International Conference on Mechanical and Aerospace Engineering (ICMAE)*. IEEE, 100-105.

Hoogeveen, JA, Steef L van de Velde, Bart Veltman. 1994. Complexity of scheduling multiprocessor tasks with prespecified processor allocations. *Discrete Applied Mathematics*, 55 (3), 259-272.

Hou, Wenjuan, Tao Fang, Zhi Pei, Qiao-Chu He. 2021. Integrated design of unmanned aerial mobility network: A data-driven risk-averse approach. *International Journal of Production Economics*, 236 108131.

Jiang, Nan, Weijun Xie. 2023. Distributionally favorable optimization: A framework for data-driven decision-making with endogenous outliers. *SIAM Journal on Optimization*, .

Kleinbekman, Imke C, Mihaela A Mitici, Peng Wei. 2018. evtol arrival sequencing and scheduling for on-demand urban air mobility. *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*. IEEE, 1-7.

Kochenderfer, Mykel J, Jessica E Holland, James P Chryssanthacopoulos. 2012. Next-generation airborne collision avoidance system. Tech. rep., Massachusetts Institute of Technology-Lincoln Laboratory Lexington United States.

Kong, Nan, Andrew J Schaefer, Shabbir Ahmed. 2013. Totally unimodular stochastic programs. *Mathematical Programming*, 138 (1-2), 1-13.

Laporte, Gilbert, François V Louveaux. 1993. The integer l-shaped method for stochastic integer programs with complete recourse. *Operations research letters*, 13 (3), 133-142.

Lei, Xiao, Siqian Shen, Yongjia Song. 2018. Stochastic maximum flow interdiction problems under heterogeneous risk preferences. *Computers & Operations Research*, 90 97-109.

Lineberger, Robin, Aijaz Hussain, Vincent Rutgers. 2018. Change is in the air: The elevated future of mobility: What's next on the horizon?, Available at `https://www2.deloitte.com/content/dam/Deloitte/us/Documents/energy-resources/di-the-elevated-future-of-mobility.pdf` (last accessed date: October 27, 2022).

Lulli, Guglielmo, Amedeo Odoni. 2007. The european air traffic flow management problem. *Transportation science*, 41 (4), 431-443.

Moug, Kati, Huiwen Jia, Siqian Shen. 2022. A shared mobility based framework for evacuation planning and operations under forecast uncertainty. *IISE Transactions*, (just-accepted), 1-33.

Mueller, Eric R, Parmial H Kopardekar, Kenneth H Goodrich. 2017. Enabling airspace integration for high-density on-demand mobility operations. *17th AIAA Aviation Technology, Integration, and Operations Conference*. 3086.

Mukherjee, Avijit, Mark Hansen. 2009. A dynamic rerouting model for air traffic flow management. *Transportation Research Part B: Methodological*, 43 (1), 159-171.

Pelegrín, Mercedes, Claudia d'Ambrosio, Rémi Delmas, Youssef Hamadi. 2021. Urban air mobility: From complex tactical conflict resolution to network design and fairness insights, .

Powell, Warren B, Huseyin Topaloglu. 2003. Stochastic programming in transportation and logistics. *Handbooks in operations research and management science*, 10 555-635.

Price, George, Douglas Helton, Kyle Jenkins, Mike Kvicala, Steve Parker, Russell Wolfe. 2020. Urban air mobility operational concept (opscon) passenger-carrying operations, Available at `https://ntrs.nasa.gov/citations/20205001587` (last accessed date: October 27, 2022).

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

Rahmaniani, Ragheb, Teodor Gabriel Crainic, Michel Gendreau, Walter Rei. 2017. The benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259 (3), 801-817.

Rajendran, Suchithra, Sharan Srinivas. 2020. Air taxi service for urban mobility: A critical review of recent developments, future challenges, and opportunities. *Transportation research part E: logistics and transportation review*, 143 102090.

Rajendran, Suchithra, Joshua Zack. 2019. Insights on strategic air taxi network infrastructure locations using an iterative constrained clustering approach. *Transportation Research Part E: Logistics and Transportation Review*, 128 470-505.

Rath, Srushti, Joseph YJ Chow. 2022. Air taxi skyport location problem with single-allocation choice-constrained elastic demand for airport access. *Journal of Air Transport Management*, 105 102294.

Rockafellar, R Tyrrell, Stanislav Uryasev, et al. 2000. Optimization of conditional value-at-risk. *Journal of risk*, 2 21-42.

Rodionova, Olga, Heather Arneson, Banavar Sridhar, Antony Evans. 2017. Efficient trajectory options allocation for the collaborative trajectory options program. *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. IEEE, 1-10.

Schrank, David, Luke Albert, Bill Eisele, Tim Lomax, et al. 2021. Urban mobility report 2021, .

Shapiro, Alexander, Shabbir Ahmed. 2004. On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14 (4), 1237-1249.

Shapiro, Alexander, Darinka Dentcheva, Andrzej Ruszczynski. 2021. *Lectures on stochastic programming: modeling and theory*. SIAM.

Shehadeh, Karmel S. 2022. Distributionally robust optimization approaches for a stochastic mobile facility fleet sizing, routing, and scheduling problem. *Transportation Science*, .

Speranza, M Grazia. 2018. Trends in transportation and logistics. *European Journal of Operational Research*, 264 (3), 830-836.

Sun, Luying, Weijun Xie, Tim Witten. 2022. Distributionally robust fair transit resource allocation during a pandemic. *Transportation science*, .

Thipphavong, David P, Rafael Apaza, Bryan Barmore, Vernol Battiste, Barbara Burian, Quang Dao, Michael Feary, Susie Go, Kenneth H Goodrich, Jeffrey Homola, et al. 2018. Urban air mobility airspace integration concepts and considerations. *2018 Aviation Technology, Integration, and Operations Conference*. 3676.

Toth, Paolo, Daniele Vigo. 2002. *The vehicle routing problem*. SIAM.

Toumazis, Iakovos, Changhyun Kwon. 2013. Routing hazardous materials on time-dependent networks using conditional value-at-risk. *Transportation Research Part C: Emerging Technologies*, 37 73-92.

United Nations. 2018. *World Urbanization Prospects 2018*. Available at `https://population.un.org/wup/Download/` (last accessed date: October 27, 2022).

Vascik, Parker D, R John Hansman. 2017. Evaluation of key operational constraints affecting on-demand mobility for aviation in the los angeles basin: ground infrastructure, air traffic control and noise. *17th AIAA Aviation Technology, Integration, and Operations Conference*. 3084.

Wang, Kai, Alexandre Jacquillat. 2020. A stochastic integer programming approach to air traffic scheduling and operations. *Operations Research*, 68 (5), 1375-1402.

Wang, Zhengyi, Daniel Delahaye, Jean-Loup Farges, Sameer Alam. 2022. Complexity optimal air traffic assignment in multi-layer transport network for urban air mobility operations. *Transportation Research Part C: Emerging Technologies*, 142 103776.

Wu, Pengcheng, Junfei Xie, Yanchao Liu, Jun Chen. 2022. Risk-bounded and fairness-aware path planning for urban air mobility operations under uncertainty. *Aerospace Science and Technology*, 127 107738.

Wu, Zhiqiang, Yu Zhang. 2021. Integrated network design and demand forecast for on-demand urban air mobility. *Engineering*, 7 (4), 473-487.

Yu, Xian, Siqian Shen, Shabbir Ahmed. 2021a. On the value of multistage stochastic facility location with risk aversion. *arXiv preprint arXiv:2105.11005*, .

Yu, Xian, Siqian Shen, Huizhu Wang. 2021b. Integrated vehicle routing and service scheduling under time and cancellation uncertainties with application in nonemergency medical transportation. *Service Science*, 13 (3), 172-191.

Zhu, Guodong, Peng Wei, Robert Hoffman, Bert Hackney. 2018a. Centralized disaggregate stochastic allocation models for collaborative trajectory options program (ctop). *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*. IEEE, 1-10.

Zhu, Guodong, Peng Wei, Robert Hoffman, Bert Hackney. 2018b. Saturation technique for optimizing planned acceptance rates in traffic management initiatives. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 3536-3543.

Appendix A. Proofs

A.1 Proof Proposition 1

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

Partition Problem: Consider $|\mathcal{I}_1|$ positive numbers $\{m_i\}_{i\in\mathcal{I}_1} \subseteq \mathbb{Z}_{++}$ having an even sum, is there a partition S_1, S_2 such that $\sum_{i\in S_1} m_i = \sum_{i\in S_2} m_i = \bar{D}, S_1 \cap S_2 = \emptyset, S_1 \cup S_2 = \mathcal{I}_1$?

We show that finding a feasible solution for a special case of FairUAM (1) can be reduced to the partition problem. We first let $\boldsymbol{p} = \mathbf{1}, \boldsymbol{q} = \mathbf{0}, \boldsymbol{c} = \mathbf{0}, \bar{L} = 0, \bar{\boldsymbol{R}} = \mathbf{0}, \boldsymbol{C} = \mathcal{M}\mathbf{1}$ where $\mathcal{M} = |\mathcal{I}_1|$. Suppose that $\epsilon = 1$ in the risk measure, i.e., $\rho(\cdot) = \mathbb{E}[\cdot]$, and the passenger demand is $\bar{D} = 1/2 \sum_{i\in\mathcal{I}_1} m_i$ for both routes. Then FairUAM (1) reduces to

$$z_1^* = \min_{\boldsymbol{\delta},\boldsymbol{y},\boldsymbol{s}} \quad \frac{1}{|\mathcal{I}_1|} \sum_{r\in[2]} s^r, \tag{20a}$$

$$\text{s.t.} \quad \sum_{r\in[2]} y_i^r \le 1, \quad \forall i \in \mathcal{I}_1, \tag{20b}$$

$$s^r + \sum_{i\in\mathcal{I}_1} m_i y_i^r \ge \bar{D}, \quad \forall r \in [2], \tag{20c}$$

$$y_i^r \in \{0,1\}, s^r \ge 0, \forall r \in [2], \forall i \in \mathcal{I}_1, \tag{20d}$$

where $[n] := \{1, 2, \ldots, n\}$. Thus, we claim that the optimal value of formulation (20) $z_1^* \le 0$ if and only if there exists a feasible solution $(\boldsymbol{s}, \boldsymbol{y})$ satisfying $\sum_{i\in\mathcal{I}_1} m_i y_i^r \ge \bar{D}, \forall r \in [2]$. This is because when $2\bar{D} = \sum_{i\in\mathcal{I}_1} m_i$, we must have $\sum_{i\in\mathcal{I}_1} m_i y_i^r = \bar{D}, \forall r \in [2]$ and $\sum_{r\in[2]} y_i^r = 1, \forall i \in \mathcal{I}_1$. Thus, solving formulation (20) is equivalent to solve the following problem

$$\sum_{r\in[2]} y_i^r = 1, \sum_{i\in\mathcal{I}_1} m_i y_i^r = \bar{D}, y_i^r \in \{0,1\}, \forall i \in \mathcal{I}_1, r \in [2],$$

which is exactly equivalent to the partition problem. Thus, checking the special case (20) of FairUAM (1) having the optimal value being equal to 0 or not is equivalent to solving the partition problem. This proves the NP-hardness of FairUAM (1). □

A.2 Proof of Proposition 2

Proof. Let us consider a special case of the second-stage problem (2). Suppose the given first stage decision $(\boldsymbol{y}, \boldsymbol{\delta})$ is feasible, and we let $\boldsymbol{p} = \mathbf{0}, \boldsymbol{q} = \mathbf{0}, \boldsymbol{c} = \mathbf{0}, \boldsymbol{\tau} = \mathbf{0}, \bar{L} = |\mathcal{T}|, \bar{\boldsymbol{R}} = |\mathcal{T}|, |\mathcal{R}| = 1$. Since the random parameters have one realization, and we let airspace capacity $\boldsymbol{C} = \mathbf{1}$ and the ground delay and air delay costs to be all one, the second-stage problem (2) reduces to

$$\min_{\boldsymbol{g},\boldsymbol{a},\boldsymbol{s},\boldsymbol{B}} \frac{1}{|\mathcal{I}_1|} \sum_{i\in\mathcal{I}_1} \left(g_i^1 + \sum_{k\in\mathcal{K}} a_{ik}^1 \right), \tag{21a}$$

$$\text{s.t.} \quad \sum_{t\in\mathcal{T}} B_{it}^k \le 1, \quad \forall i\in\mathcal{I}_1, k\in\mathcal{K}, \tag{21b}$$

$$\sum_{t\in\mathcal{T}} tB_{it}^k = 1 + g_i^1 + \sum_{\substack{id(k)\ge id(k')\\ k'\in\mathcal{K}}} a_{ik'}^1, \quad \forall i\in\mathcal{I}_1, k\in\mathcal{K}, \tag{21c}$$

$$\sum_{i\in\mathcal{I}_1} B_{it}^k \le 1, \quad \forall k\in\mathcal{K}, t\in\mathcal{T}, \tag{21d}$$

$$B_{it}^k \in \{0,1\}, \quad \forall i\in\mathcal{I}_1, k\in\mathcal{K}, t\in\mathcal{T}, \tag{21e}$$

$$g_i^1, a_{ik}^1 \in \mathbb{Z}_+, \quad \forall i\in\mathcal{I}_1, k\in\mathcal{K}. \tag{21f}$$

We show that job-shop scheduling problem (JSP) (Graham 1966), which is NP-complete even with unit processing time (see, e.g., theorem 3.5 in Hoogeveen et al. 1994), reduces to formulation (21).

Job-shop scheduling problem (JSP): There are a set $\mathcal{K}$ of machines and a set $\mathcal{I}$ of jobs. Each job $i\in\mathcal{I}$ consists of an ordered collection of operations $\{k_1,\dots,k_{|\mathcal{K}|}\}=\mathcal{K}$, and each operation must be completed at a specific machine and requires a unit of processing time on that machine. Each machine can only process one operation at a time. The goal is to optimize the order of operations completed on each machine to minimize the overall completion time (i.e., the overall delay).

In formulation (21), we let each aircraft $k\in\mathcal{K}$ represent a job and each bottleneck point $i\in\mathcal{I}_1$ denote a machine. We also let the visiting sequence of bottleneck points (i.e., the variable $\{B_{it}^k\}_{i\in\mathcal{I}_1,t\in\mathcal{T}}$) of an aircraft $k\in\mathcal{K}$ represent the ordered collection of operations of a job. Furthermore, the unit processing time of an operation is associated with the time that an aircraft passes the bottleneck point. The requirement that no two operations can be performed simultaneously is equivalent to the unit bottleneck point capacity. Particularly, the objective (21a) minimizes the overall job completion time. Constraints (21b) implies that at most one completion time for each job at each machine. Constraints (21c) specify when each job will be completed at each machine. Constraints (21d) postulate that at a given time and for a given machine, at most one job can be operated. In this case, the job-shop problem with unit processing time exactly reduces to formulation (21). This proves the NP-hardness of the second-stage problem (2). □