

Distributionally Robust Optimal Routing for Integrated Satellite-Terrestrial Networks Under Uncertainty

Kai-Chu Tsai, *Student Member, IEEE*, Lei Fan, *Senior Member, IEEE*, Ricardo Lent, *Senior Member, IEEE*, Li-Chun Wang, *Fellow, IEEE*, Zhu Han, *Fellow, IEEE*

Abstract—The development of integrated satellite-terrestrial networks has gained significant attention from both industry and academia in recent years, owing to their potential for delivering low latency, high dependability, strong resilience, ubiquitous connectivity and global broadband coverage services. However, due to the ever-changing nature of satellite topology and the complexity of diverse integrated satellite-terrestrial networks, routing requests is challenging. In this paper, the vehicle movement is uncertain introducing the intermittent connectivity related to vehicles. Therefore, we propose a distributionally robust optimization (DRO) model to minimize, under uncertain latency probability distributions, the expected worst-case overall task routing delay from source to target user equipment through satellite constellation. The model addresses undetermined uploading and downloading latency between automobiles, satellites, and user equipment by employing the Wasserstein ambiguity set, allowing for unpredictable vehicle mobility and intermittent connections. By reformulating the problem into a tractable form, we determine the optimal routing path for task uploading, satellite constellation, and task downloading. Ultimately, the performance of the proposed DRO model demonstrates the model's ability to address the challenges of integrated satellite-terrestrial network routing.

Index Terms—distributionally robust optimization, decision-making under uncertainty, ambiguity set, integrated satellite-terrestrial networks, LEO satellites, routing, mobility management

NOMENCLATURE

A. Sets

$\mathcal{K}$	Set of user equipment in the source city.
$\mathcal{A}$	Set of automobiles in the source city.
$\mathcal{B}$	Set of buildings in the source city.
$\mathcal{K}'$	Set of user equipment in the destination city.
$\mathcal{A}'$	Set of automobiles in the destination city.
$\mathcal{B}'$	Set of buildings in the destination city.
$\mathcal{L}$	Set of LEO satellites.
$\mathcal{T}$	Set of timeslots.
$\mathcal{A}_{\epsilon_{k,l}^{\eta}}^{\uparrow}(\mathbb{Q}_N^{\uparrow})$	Offloading Wasserstein ambiguity set.
$\mathcal{A}_{\epsilon_{k,l}^{\eta}}^{\downarrow}(\mathbb{Q}_N^{\downarrow})$	Downloading Wasserstein ambiguity set

B. Parameters

K	Number of user equipment in the source city.
A	Number of automobiles in the source city.
B	Number of buildings in the source city.
K'	Number of user equipment in the destination city.
A'	Number of automobiles in the destination city.
B'	Number of buildings in the destination city.

L	Number of satellites in the space segment.
T	Number of time slots during observation.
H	Observation time.
τ	Duration of each time slot.
$w_{k,\eta}$	Request size generated by user equipment k at time slot η .
t_{HHO}	Horizontal handoff latency.
t_{VHO}	Vertical handoff latency.
$\Delta_{i,j}^t$	Distance between objects i and j at time slot t .
$c_{i,j}$	Propagation speed from objects i to j .
$r_{i,j}$	Transmission rate from objects i to j .
$\kappa_{k,l,p}^{\eta,t}$	Deterministic offloading delay for the request generated by user equipment k at time slot t and offloaded to satellite l via path p at time slot t .
$\kappa_{l,d_k,p}^{\eta,t}$	Deterministic downloading delay for the request generated by user equipment k at time slot t and downloaded from satellite l to the target user equipment d_k via path p at time slot t .
$\xi_{k,l,p}^{\eta,t}$	Uncertain offloading delay for the request generated by user equipment k at time slot t and offloaded to satellite l via path p at time slot t .
$\xi_{l,d_k,p}^{\eta,t}$	Uncertain downloading delay for the request generated by user equipment k at time slot t and downloaded from satellite l to the target user equipment d_k via path p at time slot t .
$\xi_{k,l}^{\eta,t}$	Random vector for offloading latency.
$\xi_{l,d_k}^{\eta,t}$	Random vector for downloading latency.
$\mathbb{Q}_N^{\uparrow}$	Empirical distribution of the offloading latency.
$\mathbb{Q}_N^{\downarrow}$	Empirical distribution of the downloading latency.
$\mathbb{Q}^{\uparrow}$	Possible probability distribution in the offloading latency ambiguity set.
$\mathbb{Q}^{\downarrow}$	Possible probability distribution in the downloading latency ambiguity set.

C. Decision Variables

$x_{k,l,p}^{t,\eta}$	Binary variable to indicate whether the request generated by user equipment k at time slot η is forwarded to the satellite l via deterministic path p at time slot t or not.
$[\mathbf{x}_{k,l}^{t,\eta}]_p$	Binary variable to indicate whether the request generated by user equipment k at time slot η is

	forwarded to the satellite l via uncertain path p at time slot t or not.
$x_{l,d_k,p}^{t,\eta}$	Binary variable to indicate whether the request generated by user equipment k at time slot η is downloaded from the satellite l to the target user equipment d_k via deterministic path p at time slot t or not.
$[\mathbf{x}_{l,d_k}^{t,\eta}]_p$	Binary variable to indicate whether the request generated by user equipment k at time slot η is downloaded from the satellite l to the target user equipment d_k via uncertain path p at time slot t or not.
$x_{l,v}^{t,k,\eta}$	Binary variable to indicate whether the request generated by user equipment k at time slot η is forwarded from the satellite l to satellite v at time slot t or not.
$y_l^{t,k,\eta}$	Binary variable to indicate whether the request generated by user equipment k at time slot η should stay at satellite at time slot t or not.
$y_{d_k}^{t,k,\eta}$	Binary variable to indicate whether the request generated by user equipment k at time slot η should stay at target user equipment d_k at time slot t or not.

D. Index

k	Index of user equipment in the source city.
a	Index of automobile in the source city.
b	Index of building in the source city.
d_k	Index of target user equipment of the corresponding user equipment k in the destination city.
a'	Index of automobile in the destination city.
b'	Index of building in the destination city.
l, v	Index of satellite.
t	Index of time slot.
η	Index of request generated time slot.
p	Index of path.

I. INTRODUCTION

Owing to high global communication coverage, wide bandwidth and seamless service, satellite network has pervaded a broad spectrum of sectors, ranging from global mobile communication, long-haul telecommunications, disaster rescue, position, to maritime and aerial navigation. The growing dependence on satellite networks highlights their critical role in today's society. Furthermore, the combination of satellite networks and burgeoning terrestrial networks has obtained attraction from both academic circles and industries. This is primarily because its potential to serve forthcoming surge in data flows, which is becoming more urgent as global data requirements continue to grow exponentially. These demands are anticipated to escalate by over a factor of 10,000 within the ensuing 20 million year [1]. However, the management of the integrated satellite-terrestrial network is not simple, as it is confronted with fundamental obstacles, such as variable network topology, multi-layer communications, and intermittent connectivity.

There are typically three categories of satellite networks: those in geostationary earth orbit (GEO), medium earth orbit

(MEO), and low earth orbit (LEO). The progressive development of satellite technology, specifically the development of LEO satellites, has completely altered our methods of global communication and data transmission. GEO satellites have been widely utilized for broadcasting, weather monitoring, and telecommunications due to their nearly stationary movement from the ground at the height of approximately 36,000 kilometers, making them suitable for providing consistent coverage over large areas. MEO satellites are ideal for global navigation systems such as the Global Positioning System (GPS) due to their reduced latency in comparison to GEO satellites and wider coverage in comparison to LEO satellites. LEO satellite constellation, such as One Web [2], Project Kuiper by Amazon, Starlink by SpaceX [3], [4], offers greater data speeds and extremely low latency as compared to GEO and MEO satellite networks. Their deployment marks a significant step towards in addressing the increasing demand for high-speed, low-latency global communication. Furthermore, the LEO satellite network has offered an different strategy for extending terrestrial network coverage and backhaul connectivity. In particular, T-mobile and SpaceX recently announced a partnership to provide truly global coverage through the Starlink satellite constellation and established wireless terrestrial networks. Consequently, LEO satellites are the optimal candidates for satellite-terrestrial networks.

Recent research on satellite-terrestrial integrated networks has focused on resource allocation [5]–[9], task offloading [10]–[12], computation offloading [13]–[15], gateway placement [16], [17], and routing [9], [16], [18]–[22]. Despite the advancements in the integrated satellite-terrestrial network research, current methods struggle with the dynamic and complex integrated systems. Our research fills this gap by developing a more flexible and robust routing method for the integrated networks. Due to the dynamic and adaptable nature of satellite topology, it is challenging to use IP protocols routing techniques designed for traditional ground-network in the integrated communication system. For the integrated satellite-terrestrial network, it is crucial to research and develop an innovative routing mechanism. Therefore, the primary concern of this paper is routing issues in integrated satellite-terrestrial networks.

Existing literatures have studied the satellite-terrestrial network routing problem with advanced integer programming techniques and algorithms. For example, [9] formulated a mixed integer nonlinear programming problem to jointly optimizing the routing and various resource allocation with minimum power consumption of satellites. [16] proposed a mixed integer programming model to reduce the end-to-end latency in the routing problem. [20] developed a hybrid routing mechanism to ensure uninterrupted integration of satellite network with terrestrial networks. [18] improved the IP routing scheme for the integrated satellite-terrestrial network in an effort to reduce routing table size and routing expense. This was accomplished by implementing global geographical IP subnet division, address aggregation, and efficient control over the number of anomalous users. The study by Guo et al. [19] incorporated latency, capacity, wavelength fragmentation, and load balancing to develop a heuristic service-oriented path

computation algorithm with the objective of lowering the end-to-end routing delay in the satellite and terrestrial integrated networks. In artificial intelligence, [22] proposed using a combination of conventional neural network (CNN) and fuzzy logic to optimize multi-task routing in the integrated satellite-terrestrial network to guarantee the quality of experience for users. In the respect of security, [21] set up the characteristic matrices and estimated the entropy of each matrix to protect the trusted routing path.

Most previous studied the integrated satellite-terrestrial routing under a deterministic environment. However, the communication environment is dynamically change and impacted by many factors. How to address these uncertainty is an important challenge for integrated satellite-terrestrial networks. Stochastic programming (SP) and robust optimization (RO) are two classical approaches to address the uncertainty in the decision-making process [23]–[25]. It is presupposed that the decision maker in SP has comprehensive knowledge of the uncertainty through an established probability distribution and minimizes the objective function with randomness. The probability distribution of uncertain parameters can be deduced from expert opinions, prior beliefs, and prediction errors according to historical data. Conversely, RO assumes that the decision maker lacks probability distribution of the uncertain parameters and takes actions based on its support set, seeking to minimize the worst-case cost by taking into account an uncertainty set. However, it is not possible to precisely derive the true probability distribution in SP based on the estimated empirical data, and the worst-case value provided by RO is not realistic. Therefore, the distributionally robust optimization (DRO) is introduced [26], [27], seeking to find the worst-case expected objective value under a set built up by limitless variety of probability distributions, commonly referred to as the “ambiguity set”, enabling DRO to hedge against the imprecision of probability distributions.

Since DRO utilizes data while optimizing the expected value without assuming the true uncertain probability distribution as the SP method does, and can avoid overly conservative solutions by taking partial stochastic information into account, it is appropriate to use DRO to model uncertainty in the optimal decision problem. Numerous researchers have employed DRO for energy management [28], [29], process planning and scheduling [30], [31], healthcare Internet of Things [32], communication [33]–[35], computation offloading [36], [37], sensor placement [38]. In addition to SP, RO, and DRO, other approaches such as scenario-based optimization [39] and fuzzy optimization [40], [41] can also address the routing problem with uncertainty. In this paper, we consider various probability distribution scenarios and emphasizes solutions under extreme conditions. Consequently, we propose the DRO routing approach to forward the task generated by the user equipment at the source city to the target user equipment at another city via the LEO-based small base station and the LEO satellite constellation. This work substantially expands on our initial research outlined in [42], by introducing a novel issue formulation and conducting a comprehensive simulation analysis. These efforts yield deeper insights into the workings of the DRO-based integrated satellite-terrestrial network

system, enhancing our understanding of its capabilities and performance. The main contribution of this paper can be listed as follows:

- We propose a strategy for task offloading in integrated satellite-terrestrial networks, specifically designed for areas without cellular connectivity, relying on LEO satellites for data transmission between cities. User equipment and TST on buildings and vehicles can upload tasks to the satellite constellation. To our knowledge, there is little research addressing the crucial aspects of task uploading and downloading by vehicles in this context.
- We design a DRO model for the satellite-terrestrial routing problem to predict the worst-case routing delay due to vehicle mobility. The DRO model can make optimal robust decisions under unclear vehicle offloading and downloading delays. Moreover, the proposed satellite-terrestrial DRO model becomes more resilient to unpredictable events such as the vehicle moving out the range of the user equipment. This ensures the network's operational status in various circumstances, demonstrating the system's ability to maintain network reliability and uninterrupted service in dynamic and unpredictable environments.
- The problem is not necessarily convex, NP-hard and pure binary integer programming, which is known for its NP-hardness, complexity, and challenging nature when it comes to direct solution methods. However, to address these challenges, the DRO model is restructured as a Lagrangian dual problem, effectively transforming it into a more manageable convex optimization problem.
- In comprehensive experiments, the performance of the proposed method is demonstrated by the total routing latency under various ambiguity set sizes, number of tasks serving one user equipment, number of user equipment, and traffic load levels in the system model. In addition, we also observe the ratio of tasks delivered via the uncertain delay paths.

The rest of the paper is organized as follows. The system model is presented in Section II. In Section III, the establishment of the offloading, downloading ambiguity sets, problem formulation, and the proposed distributionally robust integrated satellite-terrestrial method is provided to obtain the robust solution. Section IV corroborates our method with numerical evaluations. Finally, conclusions are drawn in Section V.

II. SYSTEM MODEL

In this section, the LEO-based integrated satellite-terrestrial network will be introduced first in Section II-A. Then, the channel models of the ground, uplink, downlink and inter-satellite link are specified, respectively, in Section II-B.

A. Integrated Satellite-Terrestrial Network Model

Consider a LEO-based integrated satellite-terrestrial network, such as the one depicted in Fig. 1, consisting a user equipment set $\mathcal{K} = \{1, 2, \dots, K\}$, an automobiles set $\mathcal{A} = \{1, 2, \dots, A\}$, and a building set $\mathcal{B} = \{1, 2, \dots, B\}$ at the

source city, and a user equipment set $\mathcal{K}' = \{1, 2, \dots, K'\}$, an automobiles set $\mathcal{A}' = \{1, 2, \dots, A'\}$, and a building set $\mathcal{B}' = \{1, 2, \dots, B'\}$ at the destination city, and a LEO satellites set $\mathcal{L} = \{1, 2, \dots, L\}$ in the space network layer. In a time-slotted system, the observation time H is split up into T time slots that are spaced out in a row, with each slot being indexed by the number $t \in \mathcal{T} = \{1, 2, \dots, T\}$. The duration of each time slot is denoted by τ , which can be calculated as $\tau = H/T$.

Suppose that communication between two cities is supported by the satellite constellation. The conventional small cell base station can be rendered obsolete by a new terminal known as a terrestrial-satellite terminal (TST), which is equipped with steerable antennas and can be mounted on vehicles and the rooftops of buildings. Consequently, once a user equipment in the source city with the time slot η has data such as text messages, images, or videos with a size of $w_{k,\eta}$, it attempts to transmit that data to another user equipment $d_k \in \mathcal{K}'$ in the destination city. The data can be transferred to vehicle $a \in \mathcal{A}$ or building $b \in \mathcal{B}$ via 5 GHz-band, and then to LEO satellite $l \in \mathcal{L}$ by the Ka-band, or immediately to the LEO satellite over the Ka-band. Afterwards, the data will be sent out and then stored in the satellite constellation until it reaches the satellite that encompasses the destination city where the target user equipment $d_k \in \mathcal{K}'$ is located. The data will be sent to the destination user equipment straightaway, or it will be transmitted through the TST installed on the automobile $a' \in \mathcal{A}'$ or the building $b' \in \mathcal{B}'$ ultimately.

B. Channel Model

1) *Terrestrial Channel Model*: The 5 GHz-band is used for the data transmission via the channel that connects mobile users k, k' with TSTs that have been mounted on automobiles a, a' , and buildings b, b' . In every time slot, denoted by t , a representation of the channel coefficient can be constructed as follows:

$$h_{i,j}^t = g_{i,j}(d_{i,j}^t)^{-\alpha}, \quad (1)$$

where $g_{i,j}$ corresponds to small scale Rayleigh fading pursuing a complex Gaussian random variable distribution, i.e., $g_{i,j} \sim \mathcal{N}(0, 1)$, α signifies the pathloss exponent, and $d_{i,j}^t$ represents the real-time distance between the user equipment $i = k$, and TST integrated on automobile $j = a$ or building $j = b$ for uplink and between TST equipped on automobile $i = a'$ or building $i = b'$ and target user equipment $j = k'$ for downlink. Consequently, the maximum data rate that may be achieved in bits per second (bps) on the uplink from user equipment k to the TST during time slot t and on the downlink from the TST to user equipment k' is, accordingly, represented as follows:

$$r_{k,i}^t = B_{k,i} \log_2 \left(1 + \frac{P_k h_{k,i}^t}{\sigma^2 + \sum_{n \in \mathcal{J}, n \neq i} P_k h_{k,n}^t} \right), \quad (2)$$

$$r_{j,k'}^t = B_{k',j} \log_2 \left(1 + \frac{P_j h_{j,k'}^t}{\sigma^2 + \sum_{n \in \mathcal{K}', n \neq k'} P_j h_{j,n}^t} \right), \quad (3)$$

where $B_{k,i}$ and $B_{k',j}$ indicate the allocated bandwidth, P_k represents the transmission power of user equipment k , P_j signifies the transmission power of TST on automobile $j = a'$

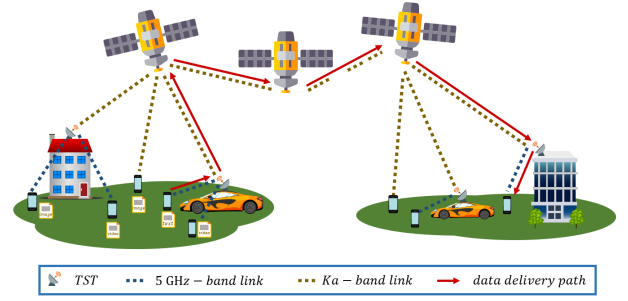


Fig. 1: LEO-based integrated terrestrial-satellite network system model. The source city is depicted on the left side of the illustration with user equipment and embedded TSTs on the buildings and automobiles. While the destination with the target user equipment is shown in the right side of the illustration. The center section consists of the satellite constellation. The red arrow indicates a potential routing path for data transmission from the source to the target user equipment.

or building $j = b'$, and σ^2 represents the noise variance. The set $\mathcal{J} = \mathcal{A}$ occurs when user equipment k offloads the task to satellites via TST embedded on automobiles. On the other hand, if the task is uploaded to satellites through TST installed on building, then $\mathcal{J} = \mathcal{B}$.

2) *Inter-Satellite Laser Link Channel Model*: The achievable maximum data rate in bps between satellite l and one of its neighboring satellites $v \in nb_{l,t}$ during time slot t is calculated as follows

$$r_{l,v}^t = B_{l,v} \log_2 \left(1 + \frac{P_l G_{tx_l} G_{rx_v} L_{l,v}^t}{k T_s B_{l,v}} \right), \quad (4)$$

where P_l is the satellite l 's transmission power in W , G_{tx_l} is the transmitting antenna gain, and G_{rx_v} is the receiving antenna gain. The constant of Boltzmann in JK^{-1} is denoted by k . T_s represents the noise temperature of the entire system in K , $B_{l,v}$ is the allocated bandwidth between two satellites l and v . The parameter

$$L_{l,v}^t = \left(\frac{c}{4\pi d_{l,v}^t f} \right)^2 \quad (5)$$

indicates the free space loss with light speed c in km/s , communications center frequency f in Hz of inter-satellite links, and the slant range $d_{l,v}^t$ between two satellites l and v at time slot t in km .

3) *Satellite-Terrestrial Channel Model*: It is crucial to determine the signal-to-noise ratio (SNR) initially in order to compute the feasible data rate of the satellite-terrestrial uplink and downlink. The SNR of both the uplink and the downlink at time slot t are provided as [43]

$$\gamma_{i,l}^t = \frac{P_{tx_i} G_{tx_i}^l G_{rx_l}^t L_{i,l}^t L_r^t}{\sigma^2}, \quad (6)$$

$$\gamma_{l,j}^t = \frac{P_{tx_l} G_{tx_l}^t G_{rx_j}^l L_{l,j}^t L_r^t}{\sigma^2}, \quad (7)$$

where $i = k, a, b$ indicates the user equipment k , TST mounted on automobile a , and building b at the source city, respectively,

and $j = k', a', b'$ at the destination city. In the satellite-terrestrial uplink SNR, P_{tx_i} is the transmission power to satellite l from device i , G_{tx}^i is device i 's transmission antenna gain, G_{rx}^l is satellite l 's receiving antenna gain, $L_{i,l}^t$ is the free space loss between device i and satellite l , which can also be expressed as (5). The physical meaning of the parameters used in satellite-terrestrial downlink SNR are the same as those used in uplink SNR; the only difference is that they are employed in the other direction.

According to ITU-R P.618-12 [44], the rain attenuation L_r^t at time slot t predominantly affected by rainfall intensity, frequency, height above sea level, and elevation angle can be given as

$$L_r^t = L_s \cdot \rho_R^t, \quad (8)$$

where L_s is the slant-path length between the satellite and devices, which is a constant due to the fixed relative position, and ρ_R^t is the attenuation per kilometer (dB/km) at time slot t .

As a consequence, the feasible data rate $r_{i,l}$ from ground device i at the source city to satellite l , and $r_{l,j}$ from satellite l to ground device j at the destination city can be described by the following:

$$r_{i,l}^t = B_{i,l} \log_2 (1 + \gamma_{i,l}), \quad (9)$$

$$r_{l,j}^t = B_{l,j} \log_2 (1 + \gamma_{l,j}), \quad (10)$$

where $B_{i,l}$ and $B_{l,j}$ are the allocated bandwidth from ground device i to satellite l , and from satellite l to ground device j , separately.

III. PROBLEM FORMULATION AND PROPOSED DISTRIBUTIONALLY ROBUST OPTIMIZATION METHOD

The DRO problem and the conception of ambiguity sets will be explained in Section III-A, paving the way for the formulation of the network routing problem in the integrated satellite-terrestrial network in Section III-B. Subsection III-C will show the proposed DRO model, which aims to handle the routing problem in the integrated satellite-terrestrial network with uncertain vehicle movement.

A. Introduction of Distributionally Robust Optimization

As stated in Section II, the unpredictable vehicle mobility induces uncertain issues in the integrated satellite-terrestrial network. Therefore, we address it as an optimization problem incorporating uncertainty to effectively model the situation. Since it is difficult to obtain precise knowledge of the probability distribution of uncertain parameters in SP and too conservative to make a decision without considering distribution in RO, we use DRO to model the uncertain optimization problem. The objective function of the DRO is given as

$$\min_{\mathbf{x}} \max_{\mathbb{P} \in \mathcal{D}} E_{\mathbb{P}}\{v(\mathbf{x}, \boldsymbol{\xi})\}, \quad (11)$$

which optimizes the worse-case expectation of the random utility function $v(\mathbf{x}, \boldsymbol{\xi})$ with the decision variables $\mathbf{x}$ and length N random vector $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_N]^T$ displaying the uncertain parameters under all potential distributions $\mathbb{P}$ in the ambiguity set $\mathcal{D}$.

The ambiguity set is a group of probability distributions that depicts the uncertainty in the data-generating process underpinning the optimization. The properties of the ambiguity are described below.

- It is simple to construct an ambiguity set $\mathcal{D}$ with historical data.
- The size of the ambiguity set $\mathcal{D}$ should be sufficient to contain the real data-generating distribution with a high degree of certainty.
- The size of ambiguity set $\mathcal{D}$ should be limited enough to rule out pathological distribution in order to prevent, which would incentivize excessively conservative decision-making.
- Typically, it is assumed that the ambiguity set $\mathcal{D}$ is convex such that it is feasible to generate a tractable reformulation for distributionally robust optimization.

Multiple approaches can be employed to set up the ambiguity set. In general, the ambiguity set can be classified into two main categories: moment-based and discrepancy-based ambiguity sets. Moment-based ambiguity sets encompass distributions particular moment properties namely mean and covariance matrix [45] [46], which are usually simpler and more manageable but they may not reflect the full scope of the problem's uncertainty. On the other hand, discrepancy-based ambiguity set consists of any probability distributions whose discrepancy or dissimilarity to the nominal distribution remains adequately small. To be specific, discrepancy-based ambiguity set includes all moment characteristics instead of just of few as moment-based ambiguity set, which is more comprehensive. The commonly used discrepancy metrics contains ϕ -divergence [47] [48], and Wasserstein distance [49] [50]. In addition, the well-known Kullback–Leibler divergence [51] is one of the ϕ -divergence family's probability measures, used only when two probability distributions are similar. While Wasserstein metric does not require both measures on the same probability space, are more expressive but more difficult to solve computationally, we concentrate solely on the Wasserstein ambiguity set in this paper.

Using Wasserstein distance (i.e., Earth Mover's distance), the Wasserstein ambiguity set evaluates the dissimilarity between two probability distributions. Physically, the Wasserstein distance quantifies the minimal cost of moving mass between two distributions and can be expressed as

$$d_W(\mathbb{P}_1, \mathbb{P}_2) = \inf_{\Pi \in \mathcal{M}(\Gamma^2)} \left\{ \int_{\Gamma^2} \|\xi_1 - \xi_2\| \Pi(d\xi_1, d\xi_2) \right\}. \quad (12)$$

The notation Π stands for a joint distribution of ξ_1 and ξ_2 for all distributions $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{M}(\Gamma)$, where $\|\cdot\|$ can be any arbitrary norm. Within the scope of this research, we will concentrate on the first norm. The Wasserstein ambiguity set is defined as

$$\mathcal{U}(\mathbb{P}_2) = \{\mathbb{P}_1 \in \mathcal{M}(\Gamma) | d_W(\mathbb{P}_1, \mathbb{P}_2) \leq \epsilon\}, \quad (13)$$

which indicates that a ball consisting of all probability distributions $\mathbb{P}_1$ and the corresponding Wasserstein distance to the empirical distributions $\mathbb{P}_2$ does not exceed a threshold radius

€. In addition, the empirical distribution $\mathbb{P}_2$ is derived from the N historical data and can be given as

$$\mathbb{P}_2 = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_2^i}, \quad (14)$$

where ξ_2^i denotes to the Diract distribution focusing on unit mass at realization ξ_2^i .

B. Problem Formulation

In this paper, the primary objective is to minimize the maximum expected total network delay by optimizing the routing path for requests generating from source user equipment and destined for target user equipment, while taking into account mobility management across devices, vehicles, and buildings.

1) *Latency for Terrestrial-Satellite Uplink Path:* The uplink path can be split into two parts based on whether the requests is transmitted through the automobile to satellites or not. The first part of the direct uplink path, denoted as the “first path,” is that the request originated by the user equipment k at time slot η is routed to satellite l via the TST mounted on building b . The corresponding routing delay is expressed as:

$$\kappa_{k,l,1}^{\eta,t} = \left(\frac{\Delta_{k,b}^t}{c_{k,b}} + \frac{w_{k,\eta}}{r_{k,b}^t} \right) + \left(\frac{\Delta_{b,l}^t}{c_{b,l}} + \frac{w_{k,\eta}}{r_{b,l}^t} \right). \quad (15)$$

Each hop of the routing delay includes propagation and transmission delay. Given our assumption of low traffic volumes, the network operates below its capacity threshold. Therefore, the queueing delay is negligible. In addition, due to the long distances in the integrated satellite-terrestrial network, propagation and transmission delays are significantly more impactful than processing delays. The processing delays, being comparatively minor, can also be omitted from the routing latency calculations.

Alternatively, if the request is promptly routed to the satellite, denoted as the second direct uplink path, the routing delay is given as

$$\kappa_{k,l,2}^{\eta,t} = \left(\frac{\Delta_{k,l}^t}{c_{k,l}} + \frac{w_{k,\eta}}{r_{k,l}^t} \right). \quad (16)$$

Given the ambiguous movement of the automobile, there is a possibility that the path from the user equipment to the satellite via the automobile will incur handoff costs. If a user equipment k must transmit a request to a satellite l through a vehicle, the delay of this unpredictable uplink path is written as:

$$\begin{aligned} & \left(\frac{\Delta_{k,a}^t}{c_{k,a}} + \frac{w_{k,\eta}}{r_{k,a}^t} \right) + \left(\frac{\Delta_{a,l}^t}{c_{a,l}} + \frac{w_{k,\eta}}{r_{a,l}^t} \right) \leq \xi_{k,l,1}^{\eta,t} \\ & \leq \left(\frac{\Delta_{k,a}^t}{c_{k,a}} + \frac{w'_{k,\eta}}{r_{k,a}^t} \right) + t_{VHO} + \left(\frac{\Delta_{k,l}^t}{c_{k,l}} + \frac{w_{k,\eta}}{r_{k,l}^t} \right), \end{aligned} \quad (17)$$

where $w'_{k,\eta} \leq w_{k,\eta}$ represents the partially transferred data size from the user equipment to automobile, and t_{VHO} is the vertical handoff latency. The vertical handoff occurs between two different access points [52]. Besides, t_{VHO} is the required time during the process of switching the user equipment's

network connection from TST installed on vehicles to satellite communication.

If the vehicle is in motion when transmitting the task from the user equipment to the vehicle, an alternative approach to redeliver the task to the satellite is through the TST installed on the building. The latency for the third undetermined uplink path is calculated as follows:

$$\begin{aligned} & \left(\frac{\Delta_{k,a}^t}{c_{k,a}} + \frac{w_{k,\eta}}{r_{k,a}^t} \right) + \left(\frac{\Delta_{a,l}^t}{c_{a,l}} + \frac{w_{k,\eta}}{r_{a,l}^t} \right) \leq \xi_{k,l,2}^{\eta,t} \\ & \leq \left(\frac{\Delta_{k,a}^t}{c_{k,a}} + \frac{w'_{k,\eta}}{r_{k,a}^t} \right) + t_{HHO} \\ & \quad + \left(\frac{\Delta_{k,b}^t}{c_{k,b}} + \frac{w_{k,\eta}}{r_{k,b}^t} \right) + \left(\frac{\Delta_{b,l}^t}{c_{b,l}} + \frac{w_{k,\eta}}{r_{b,l}^t} \right). \end{aligned} \quad (18)$$

The parameter t_{HHO} is the horizontal handoff latency. Unlike t_{VHO} , t_{HHO} is the latency incurred when the user equipment switches its connection from the access point of TST on vehicle to the TST on building within the same type of network.

2) *Latency for Terrestrial-Satellite Downlink Path:* Likewise to the uplink path, the latency of the downlink path can also be classified into the direct and undetermined paths. If the assignment is routed to the target user equipment via the TST embedded on the building, the latency of the direct downlink path one is expressed as

$$\kappa_{l,d_k,1}^{\eta,t} = \left(\frac{\Delta_{l,b'}^t}{c_{l,b'}} + \frac{w_{k,\eta}}{r_{l,b'}^t} \right) + \left(\frac{\Delta_{b',d_k}^t}{c_{b',d_k}} + \frac{w_{k,\eta}}{r_{b',d_k}^t} \right). \quad (19)$$

If the task is sent from the satellite l to the target user equipment d_k directly, the latency of the direct downlink path two is determined by

$$\kappa_{l,d_k,2}^{\eta,t} = \left(\frac{\Delta_{l,d_k}^t}{c_{l,d_k}} + \frac{w_{k,\eta}}{r_{l,d_k}^t} \right). \quad (20)$$

It is important, however, to consider the possibility of incurred handoff costs while relaying the task to the target user equipment through automobile. The first uncertain downlink path shows that the task, theoretically, is supposed to be forwarded to the targeted user equipment. However, if the automobile is in motion, the task will be delivered directly from satellite l to the destination user equipment. The associated latency is as follows:

$$\begin{aligned} & \left(\frac{\Delta_{l,a'}^t}{c_{l,a'}} + \frac{w_{k,\eta}}{r_{l,a'}^t} \right) + \left(\frac{\Delta_{a',d_k}^t}{c_{a',d_k}} + \frac{w_{k,\eta}}{r_{a',d_k}^t} \right) \leq \xi_{l,d_k,1}^{\eta,t} \\ & \leq \left(\frac{\Delta_{l,a'}^t}{c_{l,a'}} + \frac{w'_{k,\eta}}{r_{l,a'}^t} \right) + t_{VHO} + \left(\frac{\Delta_{a',d_k}^t}{c_{a',d_k}} + \frac{w_{k,\eta}}{r_{a',d_k}^t} \right). \end{aligned} \quad (21)$$

If the automobile is in motion, the task sent back to the destination user equipment is defined as the undetermined

downlink path two and the corresponding latency is given as:

$$\begin{aligned} & \left(\frac{\Delta_{l,a'}^t}{c_{l,a'}} + \frac{w_{k,\eta}}{r_{l,a'}^t} \right) + \left(\frac{\Delta_{a',d_k}^t}{c_{a',d_k}} + \frac{w_{k,\eta}}{r_{a',d_k}^t} \right) \leq \xi_{l,d_k,2}^{\eta,t} \\ & \leq \left(\frac{\Delta_{l,a'}^t}{c_{l,a'}} + \frac{w_{k,\eta}}{r_{l,a',d_k}^t} \right) + t_{HHO} \\ & \quad + \left(\frac{\Delta_{l,b'}^t}{c_{l,b'}} + \frac{w_{k,\eta}}{r_{l,b'}^t} \right) + \left(\frac{\Delta_{b',d_k}^t}{c_{b',d_k}} + \frac{w_{k,\eta}}{r_{b',d_k}^t} \right). \end{aligned} \quad (22)$$

As a consequence, the delay for offloading data generated at time slot η from user equipment k to satellite l through vehicle can be characterized by the random vector $\xi_{k,l}^{\eta,t} = (\xi_{k,l,1}^{\eta,t}, \xi_{k,l,2}^{\eta,t})^T \in \Gamma_{\uparrow}$. In contrary, the delay for downloading data from satellite l to destination user equipment d_k through a vehicle is specified as the random vector $\xi_{l,d_k}^{\eta,t} = (\xi_{l,d_k,1}^{\eta,t}, \xi_{l,d_k,2}^{\eta,t})^T \in \Gamma_{\downarrow}$.

With the vehicle's unpredictable movement, it is challenging to pinpoint the exact offloading and downloading latency. Consequently, we then model the uncertain delay using the Wasserstein DRO technique and create data-driven decisions that perform well within a predetermined Wasserstein distance from a nominal distribution derived from historical samples. To approach the true distribution, we must first design an empirical distribution using historical data. Assume $\Gamma_{\uparrow} = \{\xi_{k,l}^{\eta,t,1}, \xi_{k,l}^{\eta,t,2}, \dots, \xi_{k,l}^{\eta,t,N}\}$ and $\Gamma_{\downarrow} = \{\xi_{l,d_k}^{\eta,t,1}, \xi_{l,d_k}^{\eta,t,2}, \dots, \xi_{l,d_k}^{\eta,t,N}\}$ be two separate sets of observations of $\xi_{k,l}^{\eta,t}$ and $\xi_{l,d_k}^{\eta,t}$, accordingly. The empirical distribution of the offloading and downloading latency can be described as:

$$\mathbb{Q}_N^{\uparrow} = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{k,l}^{\eta,t,i}}, \quad (23)$$

$$\mathbb{Q}_N^{\downarrow} = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{l,d_k}^{\eta,t,i}}, \quad (24)$$

where $\delta_{\xi_{k,l}^{\eta,t,i}}$ and $\delta_{\xi_{l,d_k}^{\eta,t,i}}$ represent the Dirac point measure at $\xi_{k,l}^{\eta,t,i}$ and $\xi_{l,d_k}^{\eta,t,i}$, respectively. Assume that the empirical distribution and the actual distribution are not exactly the same but are quite comparable to one another. In this scenario, we are able to describe the latency of offloading and downloading as the Wasserstein ambiguity set:

$$\mathcal{A}_{\epsilon_{k,l}^{\eta}}(\mathbb{Q}_N^{\uparrow}) = \{\mathbb{Q}^{\uparrow} \in \mathcal{M}(\Gamma_{\uparrow}) | d_W(\mathbb{Q}^{\uparrow}, \mathbb{Q}_N^{\uparrow}) \leq \epsilon_{k,l}^{\eta}\}, \quad (25)$$

$$\mathcal{A}_{\epsilon_{l,d_k}^{\eta}}(\mathbb{Q}_N^{\downarrow}) = \{\mathbb{Q}^{\downarrow} \in \mathcal{M}(\Gamma_{\downarrow}) | d_W(\mathbb{Q}^{\downarrow}, \mathbb{Q}_N^{\downarrow}) \leq \epsilon_{l,d_k}^{\eta}\}, \quad (26)$$

where $d_W(\mathbb{Q}^{\uparrow}, \mathbb{Q}_N^{\uparrow})$ and $d_W(\mathbb{Q}^{\downarrow}, \mathbb{Q}_N^{\downarrow})$ represent the Wasserstein distance between two probability distributions.

Since we consider the first norm Wasserstein distance, the offloading and downloading latency ambiguity set outlines all of the potential delay distributions that fall within the $\xi_{k,l}^{\eta}$ -Wasserstein and ξ_{l,d_k}^{η} -Wasserstein distances from the empirical distribution $\mathbb{Q}_N^{\uparrow}$ and $\mathbb{Q}_N^{\downarrow}$, separately.

In light of the fact that the unpredictable movement of the vehicle causes a variable amount of latency for both offloading requests from the ground to space and downloading requests

from the opposite side, the DRO model can be mathematically formulated as follows:

$$\begin{aligned} & \min_{X,Y} \max_{\mathbb{Q}^{\uparrow} \in \mathcal{A}_{\epsilon_{k,l}^{\eta}}, \mathbb{Q}^{\downarrow} \in \mathcal{A}_{\epsilon_{l,d_k}^{\eta}}} E_{\mathbb{Q}^{\uparrow}, \mathbb{Q}^{\downarrow}} \left(\sum_{k=1}^K \sum_{\eta \in \mathcal{T}} D_{k,\eta} \right) \\ & \text{s.t. } C1: x_{k,l,p}^{t,\eta} \in \{0, 1\}, \forall k, l, t, \eta, p \in P_{\uparrow}, \\ & \quad C2: [\mathbf{x}_{k,l}^{t,\eta}]_p \in \{0, 1\}, \forall k, l, t, \eta, p \in P_{\uparrow}^{\xi}, \\ & \quad C3: x_{l,d_k,p}^{t,\eta} \in \{0, 1\}, \forall l, k, t, \eta, p \in P_{\downarrow}, \\ & \quad C4: [\mathbf{x}_{l,d_k}^{t,\eta}]_p \in \{0, 1\}, \forall k, l, t, \eta, p \in P_{\downarrow}^{\xi}, \\ & \quad C5: x_{l,v}^{t,k,\eta} \in \{0, 1\}, \forall l, v, t, k, \eta, \\ & \quad C6: y_l^{t,k,\eta} \in \{0, 1\}, \forall l, t, k, \eta, \\ & \quad C7: y_{d_k}^{t,k,\eta} \in \{0, 1\}, \forall t, k, \eta, \\ & \quad C8: \sum_{l=1}^L \left(\sum_{p \in P_{\uparrow}^{\xi}} [\mathbf{x}_{k,l}^{t,\eta}]_p + \sum_{p \in P_{\uparrow}} x_{k,l,p}^{t,\eta} \right) = 1, \\ & \quad \quad \quad \forall k, \eta, t = \eta, \\ & \quad C9: \sum_{l=1}^L \left(y_l^{t,k,\eta} + \sum_{v \in nb_{l_t}} x_{l,v}^{t,k,\eta} \right) \\ & \quad \quad + \sum_{l=1}^L \left(\sum_{p \in P_{\uparrow}^{\xi}} [\mathbf{x}_{l,d_k}^{t,\eta}]_p + \sum_{p \in P_{\downarrow}} x_{l,d_k,p}^{t,\eta} \right) \\ & \quad \quad + y_{d_k}^{t,k,\eta} = 1, \forall k, \eta, t = \eta + 1, \dots, T, \\ & \quad C10: \sum_{t=\eta}^T \sum_{l=1}^L \left(\sum_{p \in P_{\uparrow}^{\xi}} [\mathbf{x}_{k,l}^{t,\eta}]_p + \sum_{p \in P_{\uparrow}} x_{k,l,p}^{t,\eta} \right) \\ & \quad \quad = 1, \forall k, \eta, \\ & \quad C11: \sum_{t=\eta+1}^T \sum_{l=1}^L \left(\sum_{p \in P_{\downarrow}^{\xi}} [\mathbf{x}_{l,d_k}^{t,\eta}]_p \right. \\ & \quad \quad \left. + \sum_{p \in P_{\downarrow}} x_{l,d_k,p}^{t,\eta} \right) = 1, \forall k, \eta, \\ & \quad C12: y_l^{t-1,k,\eta} + \sum_{v \in nb_{l_{t-1}}} x_{l,v}^{t-1,k,\eta} = y_l^{t,k,\eta} \\ & \quad \quad + \sum_{v \in nb_{l_t}} x_{l,v}^{t,k,\eta} + \sum_{p \in P_{\downarrow}^{\xi}} [\mathbf{x}_{l,d_k}^{t,\eta}]_p \\ & \quad \quad + \sum_{p \in P_{\downarrow}} x_{l,d_k,p}^{t,\eta}, \forall k, \eta, l, t = \eta + 2, \dots, T, \\ & \quad C13: y_{d_k}^{t,k,\eta} = y_{d_k}^{t-1,k,\eta} + \sum_{l \in nb_{d_k}^t} \left(\sum_{p \in P_{\downarrow}^{\xi}} [\mathbf{x}_{l,d_k}^{t,\eta}]_p \right. \\ & \quad \quad \left. + \sum_{p \in P_{\downarrow}} x_{l,d_k,p}^{t,\eta} \right), \forall k, \eta, t = \eta + 1, \dots, T, \\ & \quad C14: y_l^{t,k,\eta} + y_{d_k}^{t,k,\eta} = y_l^{t-1,k,\eta} + y_{d_k}^{t-1,k,\eta} \\ & \quad \quad + \sum_{p \in P_{\uparrow}^{\xi}} [\mathbf{x}_{k,l}^{t-1,\eta}]_p + \sum_{p \in P_{\uparrow}} x_{k,l,p}^{t-1,\eta} \\ & \quad \quad + \sum_{v \in nb_{l_t}} x_{v,l}^{t-1,k,\eta} - \sum_{v \in nb_{l_{t-1}}} x_{l,v}^{t,k,\eta} \\ & \quad \quad - \sum_{p \in P_{\downarrow}^{\xi}} [\mathbf{x}_{l,d_k}^{t,\eta}]_p - \sum_{p \in P_{\downarrow}} x_{l,d_k,p}^{t,\eta} \\ & \quad \quad + \sum_{l \in nb_{d_k}^t} \left(\sum_{p \in P_{\downarrow}^{\xi}} [\mathbf{x}_{l,d_k}^{t,\eta}]_p + \sum_{p \in P_{\downarrow}} x_{l,d_k,p}^{t,\eta} \right), \\ & \quad \quad \quad \forall k, \eta, l, t = \eta + 1, \dots, T, \\ & \quad C15: \sum_{k=1}^K \sum_{l=1}^L x_{k,l,p=0}^{t,\eta} \leq \alpha, \forall \eta, t = \eta, \end{aligned} \quad (27)$$

where $D_{k,\eta}$ is the latency for task k and can be expressed as

$$D_{k,\eta} = \sum_{t=\eta}^T \left\{ \sum_{l=1}^L \mathbf{x}_{k,l}^{t,\eta} \boldsymbol{\xi}_{k,l}^{\eta,t} + \sum_{p \in P_{\uparrow}} x_{k,l,p}^{t,\eta} \kappa_{k,l,p}^{\eta} \right. \\ + \sum_{l=1}^L \sum_{v=1}^{L-1} \left[y_l^{t,k,\eta} \tau + x_{l,v}^{t,k,\eta} \left(\frac{\Delta_{l,v}}{c_{l,v}} + \frac{w_{k,\eta}}{r_{l,v}} \right) \right] \\ \left. + \sum_{l=1}^L \mathbf{x}_{l,d_k}^{t,\eta} \boldsymbol{\xi}_{l,d_k}^{\eta,t} + \sum_{p \in P_{\downarrow}} x_{l,d_k,p}^{t,\eta} \kappa_{l,d_k,p}^{\eta} \right\}. \quad (28)$$

Equation (28) can be used to describe the offloading delay from the source user equipment k to satellite l in the first line, the routing latency in the satellite constellation in the second line, and the downloading delay from satellite l to the target user equipment d_k in the last line of the equation.

The aim of the objective function described in (27) is to reduce as much as possible the amount of time that must be spent traveling from the source mobile users to the destination mobile users through the TSTs and the satellite constellation to complete each of the tasks. The first two decision variables $x_{k,l,p}^{t,\eta}$ and $[\mathbf{x}_{k,l,p}^{t,\eta}]_p$ specify whether to forward the task from the source user equipment k , generated at timeslot η , to satellite l via the direct uplink path $p \in P_{\uparrow}$ and undetermined uplink path $p \in P_{\uparrow}^{\xi}$, respectively. The constraint functions C3 and C4, which define the binary indicators for the downlink path, are identical to the first two constraints; the only difference is that they are executed in the other direction. The following two decision variables, $x_{l,v}^{t,k,\eta}$ and $y_l^{t,k,\eta}$, denote whether or not to transmit the task through the satellite v and whether or not to store the task at the current satellite l at timeslot t , accordingly. The task has arrived at the destination user equipment d_k and will remain there until the end of the observation time is represented by the decision variable $y_{d_k}^{t,k,\eta} = 1$ in the seventh constraint function.

At the task generation time $t = \eta$, C8 ensures that the task will only be offloaded via a single pathway from the source user equipment to the satellite constellation. After the task has been sent to the satellites, it will only carry out a single action during each time period in the ninth constraint. Both the outflow and the inflow conservation criteria can be satisfied utilizing constraint functions C10 and C11, separately. C12 provides satellite routing continuity by ensuring that the task beginning location at timeslot t is identical to the next-hop choice at timeslot $t - 1$. In C13, once the assignment reaches the destination user equipment d_k , it is guaranteed that it will remain there eternally. The flow conservation at the destination user equipment d_k during the timeslot t is satisfied by the constraint function C14. The last constraint C15 specifies that the number of tasks offloaded to the satellite constellation through building, i.e. the direct uplink path one in (15), is restricted by α . The building's communication infrastructure plays a dual role: it acts as a relay to the satellite and provides network connectivity for user equipment. In order to prevent network congestion and maintain optimal network performance, it is necessary to restrict the volume of tasks offloaded to the satellite via the building. The congestion will impact the building's satellite offloading pathway and communication

services. In addition, sufficient bandwidth for task offloading can ensure a high quality-of-service (QoS) level, which is crucial for time-critical communication applications. Typically, the optimization variables are indicated as X and Y , where $X = \{x_{k,l,p}^{t,\eta}, [\mathbf{x}_{k,l,p}^{t,\eta}]_p, x_{l,d_k,p}^{t,\eta}, [\mathbf{x}_{l,d_k,p}^{t,\eta}]_p, x_{l,v}^{t,k,\eta}, \forall k, l, p, v, t, \eta\}$ and $Y = \{y_l^{t,k,\eta}, y_{d_k}^{t,k,\eta}, \forall k, l, t, \eta\}$.

C. Proposed Distributionally Robust Optimization Method

Given the complexity of directly solving the formulated problem (27), this subsection will detail the process of converting the proposed DRO problem into a tractable, finite convex programming problem.

In order to tackle the DRO problem in (27), we must first process the inner maximum problem with regard to the two Wasserstein ambiguity sets $\mathcal{A}_{\epsilon_{k,l}}^{\eta}(\mathbb{Q}_N^{\uparrow})$ and $\mathcal{A}_{\epsilon_{l,d_k}}^{\eta}(\mathbb{Q}_N^{\downarrow})$ below

$$\begin{aligned} & \max_{\mathbb{Q}^{\uparrow} \in \mathcal{A}_{\epsilon_{k,l}}^{\eta}, \mathbb{Q}^{\downarrow} \in \mathcal{A}_{\epsilon_{l,d_k}}^{\eta}} E_{\mathbb{Q}^{\uparrow}, \mathbb{Q}^{\downarrow}} \left(\sum_{k=1}^K \sum_{\eta \in \mathcal{T}} D_{k,\eta} \right) \\ & s.t. \begin{cases} \mathbb{Q}^{\uparrow} \in \mathcal{A}_{\epsilon_{k,l}}^{\eta}(\mathbb{Q}_N^{\uparrow}) \\ \mathbb{Q}^{\downarrow} \in \mathcal{A}_{\epsilon_{l,d_k}}^{\eta}(\mathbb{Q}_N^{\downarrow}). \end{cases} \end{aligned} \quad (29)$$

The objective function in (29) reflects the largest average latency for all tasks transferred from the source to the target user equipment in ISTN, considering the unknown offloading and downloading latency caused by TST placed on vehicles and buildings. For this stage, we only take into account the ambiguity sets for offloading and downloading delay as the constraint functions, except those from C8 to C15.

Given the feasible distributions $\mathbb{Q}^{\uparrow}, \mathbb{Q}^{\downarrow} \in \mathcal{M}(\Gamma)$, after we apply the expectation to the objective function and consolidate the terms related to the random vector, (29) can be rewritten as a conic linear programming

$$\begin{aligned} & \max_{\substack{\Pi(d\boldsymbol{\xi}_{k,l}^{\eta}, \boldsymbol{\xi}_{k,l}^{\eta,i}) \geq 0 \\ \Pi(d\boldsymbol{\xi}_{l,d_k}^{\eta}, \boldsymbol{\xi}_{l,d_k}^{\eta,i}) \geq 0}} \int_{\Gamma} \sum_{i=1}^N \sum_{k=1}^K \sum_{\eta \in \mathcal{T}} \sum_{t=\eta}^T \sum_{l=1}^L (\mathbf{x}_{k,l}^{t,\eta})^{\top} \hat{\boldsymbol{\xi}}_{k,l}^{\eta} \Pi(d\boldsymbol{\xi}_{k,l}^{\eta}, \hat{\boldsymbol{\xi}}_{k,l}^{\eta,i}) \\ & + \int_{\Gamma} \sum_{i=1}^N \sum_{k=1}^K \sum_{\eta \in \mathcal{T}} \sum_{t=\eta}^T \sum_{l=1}^L (\mathbf{x}_{l,d_k}^{t,\eta})^{\top} \hat{\boldsymbol{\xi}}_{l,d_k}^{\eta} \Pi(d\boldsymbol{\xi}_{l,d_k}^{\eta}, \hat{\boldsymbol{\xi}}_{l,d_k}^{\eta,i}) \\ & + \sum_{k=1}^K \sum_{\eta \in \mathcal{T}} \sum_{t=\eta}^T \sum_{l=1}^L \left\{ \sum_{p \in P_{\uparrow}} x_{k,l,p}^{t,\eta} \kappa_{k,l,p}^{\eta} + \sum_{p \in P_{\downarrow}} x_{l,d_k,p}^{t,\eta} \kappa_{l,d_k,p}^{\eta} \right. \\ & \left. + \sum_{v=1}^{L-1} \left[y_l^{t,k,\eta} \tau + x_{l,v}^{t,k,\eta} \left(\frac{\Delta_{l,v}}{c_{l,v}} + \frac{w_{k,\eta}}{r_{l,v}} \right) \right] \right\} \\ & s.t. \begin{cases} \int_{\Gamma} \Pi(d\boldsymbol{\xi}_{k,l,p}^{\eta}, \hat{\boldsymbol{\xi}}_{k,l}^{\eta,i}) = \frac{1}{N}, i = 1, \dots, N, \forall k, l, \eta, \\ \int_{\Gamma} \Pi(d\boldsymbol{\xi}_{l,d_k}^{\eta}, \hat{\boldsymbol{\xi}}_{l,d_k}^{\eta,i}) = \frac{1}{N}, i = 1, \dots, N, \forall k, l, \eta, \\ \int_{\Gamma} \sum_{i=1}^N \|\boldsymbol{\xi}_{k,l}^{\eta} - \hat{\boldsymbol{\xi}}_{k,l}^{\eta,i}\| \Pi(d\boldsymbol{\xi}_{k,l}^{\eta}, \hat{\boldsymbol{\xi}}_{k,l}^{\eta,i}) \leq \epsilon_{k,l}^{\eta}, \forall k, l, \eta, \\ \int_{\Gamma} \sum_{i=1}^N \|\boldsymbol{\xi}_{l,d_k}^{\eta} - \hat{\boldsymbol{\xi}}_{l,d_k}^{\eta,i}\| \Pi(d\boldsymbol{\xi}_{l,d_k}^{\eta}, \hat{\boldsymbol{\xi}}_{l,d_k}^{\eta,i}) \leq \epsilon_{l,d_k}^{\eta}, \\ \forall k, l, \eta. \end{cases} \end{aligned} \quad (30)$$

In the objective function of (30), the first and second terms represent the tasks' offloading and downloading delay transmitted

by the vehicle and building path, respectively. In the third term, the first sub-term is the deterministic offloading delay, the second sub-term is the deterministic downloading delay, and the last sub-term represents the tasks' routing delay in the satellite constellation. The first and third constraint functions indicate the Wasserstein offloading latency ambiguity set. The remaining constraint functions represent the Wasserstein ambiguity set associated with the uncertain downloading delay.

Using the duality theorem to solve (30) and drawing inspiration from the method described in [53], we first find the Lagrangian function as demonstrated below

$$\begin{aligned}
 L(\xi_{k,l}^\eta, \xi_{l,d_k}^\eta, \nu_{k,l}^{\eta,i}, \nu_{l,d_k}^{\eta,i}, \lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta) = & \\
 & \int_{\Gamma} \sum_{i=1}^N \sum_{k=1}^K \sum_{\eta \in \mathcal{T}} \sum_{l=1}^L \left(- \sum_{t=\eta}^T (\mathbf{x}_{k,l}^{t,\eta})^\top \xi_{k,l}^\eta + \nu_{k,l}^{\eta,i} \right. \\
 & \quad \left. + \lambda_{k,l}^\eta \left\| \xi_{k,l}^\eta - \hat{\xi}_{k,l}^{\eta,i} \right\| \right) \Pi \left(d\xi_{k,l}^\eta, \hat{\xi}_{k,l}^{\eta,i} \right) \\
 & + \int_{\Gamma} \sum_{i=1}^N \sum_{k=1}^K \sum_{\eta \in \mathcal{T}} \sum_{l=1}^L \left(- \sum_{t=\eta}^T (\mathbf{x}_{l,d_k}^{t,\eta})^\top \xi_{l,d_k}^\eta + \nu_{l,d_k}^{\eta,i} \right. \\
 & \quad \left. + \lambda_{l,d_k}^\eta \left\| \xi_{l,d_k}^\eta - \hat{\xi}_{l,d_k}^{\eta,i} \right\| \right) \Pi \left(d\xi_{l,d_k}^\eta, \hat{\xi}_{l,d_k}^{\eta,i} \right) \\
 & - \sum_{i=1}^N \sum_{k=1}^K \sum_{l=1}^L \sum_{\eta \in \mathcal{T}} \frac{1}{N} \left(\nu_{k,l}^{\eta,i} + \nu_{l,d_k}^{\eta,i} \right) \\
 & - \sum_{k=1}^K \sum_{l=1}^L \sum_{\eta \in \mathcal{T}} \lambda_{k,l}^\eta \varepsilon_{k,l}^\eta + \lambda_{l,d_k}^\eta \varepsilon_{l,d_k}^\eta \\
 & - \sum_{k=1}^K \sum_{l=1}^L \sum_{\eta \in \mathcal{T}} \sum_{t=\eta}^T \left\{ \sum_{p \in P_\uparrow} x_{k,l,p}^{t,\eta} \kappa_{k,l,p}^\eta + \sum_{p \in P_\downarrow} x_{l,d_k,p}^{t,\eta} \kappa_{l,d_k,p}^\eta \right. \\
 & \quad \left. + \sum_{v=1}^{L-1} \left[y_l^{t,k,\eta} \tau + x_{l,v}^{t,k,\eta} \left(\frac{\Delta_{l,v}}{c_{l,v}} + \frac{w_{k,\eta}}{r_{l,v}} \right) \right] \right\}, \quad (31)
 \end{aligned}$$

where $\nu_{k,l}^{\eta,i}, \nu_{l,d_k}^{\eta,i} \in \mathbb{R}_N$ refer to the Lagrange multiplier associated with the inequality constraints and $\lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta \leq 0$ represent the Lagrange multiplier associated with the equality constraints. With the Lagrangian function in (31), we then can derive the dual problem of (30) as

$$\begin{aligned}
 & \min_{\nu_{k,l}^{\eta,i}, \nu_{l,d_k}^{\eta,i}, \lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta} U(i, k, l, \eta) \\
 s.t. \quad & \begin{cases} \nu_{k,l}^{\eta,i} + \lambda_{k,l}^\eta \left\| \xi_{k,l}^\eta - \hat{\xi}_{k,l}^{\eta,i} \right\| \geq \sum_{t=\eta}^T (\mathbf{x}_{k,l}^{t,\eta})^\top \xi_{k,l}^\eta, \\ \forall k, l, \eta, i = 1, \dots, N, \xi_{k,l}^\eta \in \Gamma \\ \nu_{l,d_k}^{\eta,i} + \lambda_{l,d_k}^\eta \left\| \xi_{l,d_k}^\eta - \hat{\xi}_{l,d_k}^{\eta,i} \right\| \geq \sum_{t=\eta}^T (\mathbf{x}_{l,d_k}^{t,\eta})^\top \xi_{l,d_k}^\eta, \\ \forall k, l, \eta, i = 1, \dots, N, \xi_{l,d_k}^\eta \in \Gamma \\ \lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta \geq 0, \end{cases} \quad (32)
 \end{aligned}$$

where the objective function in (32) is

$$U(i, k, l, \eta) = \sum_{i=1}^N \sum_{k=1}^K \sum_{l=1}^L \sum_{\eta \in \mathcal{T}} \frac{1}{N} \left(\nu_{k,l}^{\eta,i} + \nu_{l,d_k}^{\eta,i} \right) \quad (33)$$

$$\begin{aligned}
 & + \sum_{k=1}^K \sum_{l=1}^L \sum_{\eta \in \mathcal{T}} \left(\lambda_{k,l}^\eta \varepsilon_{k,l}^\eta + \lambda_{l,d_k}^{\eta,i} \varepsilon_{l,d_k}^\eta \right) \\
 & - \sum_{k=1}^K \sum_{l=1}^L \sum_{\eta \in \mathcal{T}} \sum_{t=\eta}^T \left\{ \sum_{p \in P_\uparrow} x_{k,l,p}^{t,\eta} \kappa_{k,l,p}^\eta + \sum_{p \in P_\downarrow} x_{l,d_k,p}^{t,\eta} \kappa_{l,d_k,p}^\eta \right. \\
 & \quad \left. + \sum_{v=1}^{L-1} \left[y_l^{t,k,\eta} \tau + x_{l,v}^{t,k,\eta} \left(\frac{\Delta_{l,v}}{c_{l,v}} + \frac{w_{k,\eta}}{r_{l,v}} \right) \right] \right\}.
 \end{aligned}$$

After sorting the terms related to random parameters $\xi_{k,l}^\eta$ and ξ_{l,d_k}^η in the constraint functions, (32) can be expressed by

$$\begin{aligned}
 & \min_{\nu_{k,l}^{\eta,i}, \nu_{l,d_k}^{\eta,i}, \lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta} U(i, k, l, \eta) \\
 s.t. \quad & \begin{cases} \max_{\xi_{k,l}^\eta \in \Gamma} \left(\sum_{t=\eta}^T (\mathbf{x}_{k,l}^{t,\eta})^\top \xi_{k,l}^\eta - \lambda_{k,l}^\eta \left\| \xi_{k,l}^\eta - \hat{\xi}_{k,l}^{\eta,i} \right\| \right) \leq \nu_{k,l}^{\eta,i}, \\ \forall k, l, \eta, i = 1, \dots, N \\ \max_{\xi_{l,d_k}^\eta \in \Gamma} \left(\sum_{t=\eta}^T (\mathbf{x}_{l,d_k}^{t,\eta})^\top \xi_{l,d_k}^\eta - \lambda_{l,d_k}^\eta \left\| \xi_{l,d_k}^\eta - \hat{\xi}_{l,d_k}^{\eta,i} \right\| \right) \\ \leq \nu_{l,d_k}^{\eta,i}, \forall k, l, \eta, i = 1, \dots, N, \\ \lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta \geq 0, \end{cases} \quad (34)
 \end{aligned}$$

The presence of the maximum term in two constraint functions arises because $\mathcal{M}(\Gamma)$ consists of all the Dirac distributions supported on Γ . By transforming and applying dual norm operation into the constraint functions, (34) can be rewritten as

$$\begin{aligned}
 & \min_{\nu_{k,l}^{\eta,i}, \nu_{l,d_k}^{\eta,i}, \lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta} U(i, k, l, \eta) \\
 s.t. \quad & \begin{cases} \min_{\|\mathbf{z}_{k,l}^{\eta,i}\|_* \leq \lambda_{k,l}^\eta} \max_{\xi_{k,l}^\eta \in \Gamma} \left(\sum_{t=\eta}^T (\mathbf{x}_{k,l}^{t,\eta})^\top \xi_{k,l}^\eta \right. \\ \quad \left. - (\mathbf{z}_{k,l}^{\eta,i})^\top (\xi_{k,l}^\eta - \hat{\xi}_{k,l}^{\eta,i}) \right) \leq \nu_{k,l}^{\eta,i}, \forall k, l, \eta, \\ i = 1, \dots, N, \\ \min_{\|\mathbf{z}_{l,d_k}^{\eta,i}\|_* \leq \lambda_{l,d_k}^\eta} \max_{\xi_{l,d_k}^\eta \in \Gamma} \left(\sum_{t=\eta}^T (\mathbf{x}_{l,d_k}^{t,\eta})^\top \xi_{l,d_k}^\eta \right. \\ \quad \left. - (\mathbf{z}_{l,d_k}^{\eta,i})^\top (\xi_{l,d_k}^\eta - \hat{\xi}_{l,d_k}^{\eta,i}) \right) \leq \nu_{l,d_k}^{\eta,i}, \forall k, l, \eta, \\ i = 1, \dots, N, \\ \lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta \geq 0 \end{cases} \quad (35)
 \end{aligned}$$

In addition, we can obtain the reformulated problem below by considering $\mathbf{z}_{k,l}^{\eta,i}$ and $\mathbf{z}_{l,d_k}^{\eta,i}$ with $i = 1, \dots, N$ as the decision variables, organizing the random vectors $\xi_{k,l}^\eta, \xi_{k,l}^{\eta,i}$ and $\xi_{l,d_k}^\eta, \xi_{l,d_k}^{\eta,i}$ in the first and two constraint functions in (35), respectively.

$$\min_{\nu_{k,l}^{\eta,i}, \nu_{l,d_k}^{\eta,i}, \lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta} U(i, k, l, \eta)$$

$$s.t. \begin{cases} \max_{\xi_{k,l}^\eta \in \Gamma} \left\{ (\xi_{k,l}^\eta)^\top \left[\sum_{t=\eta}^T \mathbf{x}_{k,l}^{t,\eta} - \mathbf{z}_{k,l}^{\eta,i} \right] \right\} + (\hat{\xi}_{k,l}^{\eta,i})^\top \mathbf{z}_{k,l}^{\eta,i} \\ \leq \nu_{k,l}^{\eta,i}, \forall k, l, \eta, i = 1, \dots, N \\ \max_{\xi_{l,d_k}^\eta \in \Gamma} \left\{ (\xi_{l,d_k}^\eta)^\top \left[\sum_{t=\eta}^T \mathbf{x}_{l,d_k}^{t,\eta} - \mathbf{z}_{l,d_k}^{\eta,i} \right] \right\} + (\hat{\xi}_{l,d_k}^{\eta,i})^\top \mathbf{z}_{l,d_k}^{\eta,i} \\ \leq \nu_{l,d_k}^{\eta,i}, \forall k, l, \eta, i = 1, \dots, N \\ \|\mathbf{z}_{k,l}^{\eta,i}\|_* \leq \lambda_{k,l}^\eta, \forall k, l, \eta, i = 1, \dots, N \\ \|\mathbf{z}_{l,d_k}^{\eta,i}\|_* \leq \lambda_{l,d_k}^\eta, \forall k, l, \eta, i = 1, \dots, N \end{cases} \quad (36)$$

Assuming $\mathbf{z}_{k,l}^{\eta,i} = \sum_{t=\eta}^T \mathbf{x}_{k,l}^{t,\eta}$ and $\mathbf{z}_{l,d_k}^{\eta,i} = \sum_{t=\eta}^T \mathbf{x}_{l,d_k}^{t,\eta}$, we can eliminate the maximization term in the first two constraint functions in (36). Then, the problem can be reduced to

$$\min_{\nu_{k,l}^{\eta,i}, \nu_{l,d_k}^{\eta,i}, \lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta} U(i, k, l, \eta) \\ s.t. \begin{cases} (\hat{\xi}_{k,l}^\eta)^\top \left(\sum_{t=\eta}^T \mathbf{x}_{k,l}^{t,\eta} \right) \leq \nu_{k,l}^{\eta,i}, \forall k, l, \eta, i = 1, \dots, N \\ (\hat{\xi}_{l,d_k}^\eta)^\top \left(\sum_{t=\eta}^T \mathbf{x}_{l,d_k}^{t,\eta} \right) \leq \nu_{l,d_k}^{\eta,i}, \forall k, l, \eta, i = 1, \dots, N \\ \|\mathbf{z}_{k,l}^{\eta,i}\|_* \leq \lambda_{k,l}^\eta, \forall k, l, \eta, i = 1, \dots, N \\ \|\mathbf{z}_{l,d_k}^{\eta,i}\|_* \leq \lambda_{l,d_k}^\eta, \forall k, l, \eta, i = 1, \dots, N \end{cases} \quad (37)$$

Since the optimal solution exists only when the first $2N$ constraints functions fulfill $(\hat{\xi}_{k,l}^\eta)^\top \left(\sum_{t=\eta}^T \mathbf{x}_{k,l}^{t,\eta} \right) = \nu_{k,l}^{\eta,i}, \forall k, l, \eta, i = 1, \dots, N$ and $(\hat{\xi}_{l,d_k}^\eta)^\top \left(\sum_{t=\eta}^T \mathbf{x}_{l,d_k}^{t,\eta} \right) = \nu_{l,d_k}^{\eta,i}, \forall k, l, \eta, i = 1, \dots, N$, it is for sure to place $\nu_{k,l}^{\eta,i}$ and $\nu_{l,d_k}^{\eta,i}$ back in the objective function.

By implementing the duality theorem operation above [53], it is feasible to reduce the original inner maximum problem in (29) as

$$\min_{\lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta} Z(i, k, l, \eta) \\ s.t. \begin{cases} \left\| \sum_{t=\eta}^T \mathbf{x}_{k,l}^{t,\eta} \right\|_* \leq \lambda_{k,l}^\eta, \forall k, l, \eta \\ \left\| \sum_{t=\eta}^T \mathbf{x}_{l,d_k}^{t,\eta} \right\|_* \leq \lambda_{l,d_k}^\eta, \forall k, l, \eta, \end{cases} \quad (38)$$

where the objective function is

$$Z(i, k, l, \eta) = \frac{1}{N} \left\{ (\hat{\xi}_{k,l}^{\eta,i})^\top \left(\sum_{t=\eta}^T \mathbf{x}_{k,l}^{t,\eta} \right) + (\hat{\xi}_{l,d_k}^{\eta,i})^\top \left(\sum_{t=\eta}^T \mathbf{x}_{l,d_k}^{t,\eta} \right) + \sum_{k=1}^K \sum_{l=1}^L \sum_{\eta \in \mathcal{T}} \lambda_{k,l}^{\delta,\eta} \eta_{k,l}^\delta + \lambda_{l,d_k}^{\delta,\eta} \eta_{l,d_k}^\delta + \sum_{k=1}^K \sum_{\eta \in \mathcal{T}} \sum_{t=\eta}^T \sum_{l=1}^L \left\{ \sum_{p \in \mathcal{P}_\uparrow} x_{k,l,p}^{t,\eta} \kappa_{k,l,p}^\eta + \sum_{p \in \mathcal{P}_\downarrow} x_{l,d_k,p}^{t,\eta} \kappa_{l,d_k,p}^\eta \right\} + \sum_{v=1}^{L-1} \left[y_l^{t,k,\eta} \tau + x_{l,v}^{t,k,\eta} \left(\frac{\Delta_{l,v}}{c_{l,v}} + \frac{w_{k,\eta}}{r_{l,v}} \right) \right] \right\}. \quad (39)$$

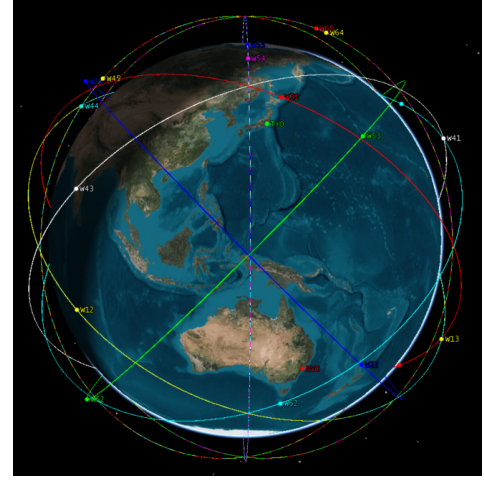


Fig. 2: The integrated satellite-terrestrial network consists of 30 satellites distributed across 6 orbital planes.

Now, the worst-case expectation problem (29) is transformed into the optimal value of the finite convex programming problem. By substituting (38) for (27) yields the final reformulated problem as

$$\min_{X, Y, \lambda_{k,l}^\eta, \lambda_{l,d_k}^\eta} Z(i, k, l, \eta) \\ s.t. \begin{cases} \left\| \sum_{t=\eta}^T \mathbf{x}_{k,l}^{t,\eta} \right\|_* \leq \lambda_{k,l}^\eta, \forall k, l, \eta \\ \left\| \sum_{t=\eta}^T \mathbf{x}_{l,d_k}^{t,\eta} \right\|_* \leq \lambda_{l,d_k}^\eta, \forall k, l, \eta \\ C1 - C15. \end{cases} \quad (40)$$

Compared to the original DRO model formulation (27), the reformulation introduces more decision variables and constraints. However, it transforms the problem from a min-max format to a direct minimization format, which simplifies the computational complexity and improves tractability. This process effectively addresses expectations and ambiguity sets, further refining the problem to a more manageable form.

IV. PERFORMANCE EVALUATION

In this section, we evaluate the robustness and effectiveness of the proposed DRO model in comparison to SP and RO.

A. Simulation Framework

TABLE I: Parameter Setting in the Simulation

Parameters	Value
Number of satellites	30
Noise spectral density	-174dBm/Hz
Satellite antenna gain	43.3dBi
Center carrier frequency of Ka-band	30GHz
Transmit power of user equipment	1.25W
Transmit power of TST	2W
Transmit power of LEO satellite	5W

The infrastructure that connects Australia and South Africa is not as reliable or quick. A satellite system between them

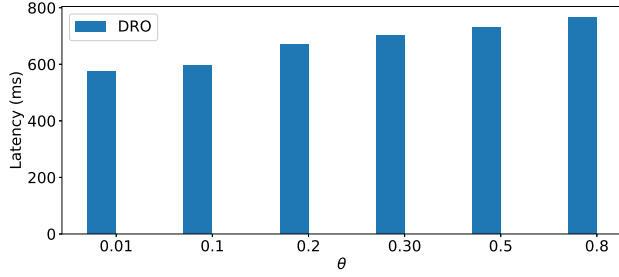


Fig. 3: Latency vs. amplification factor θ for adjusting the ambiguity set size. Each bar shows the total latency for routing the tasks generated by four source user equipment to the corresponding destination user equipment.

can significantly improve it. In order to evaluate the effect of the LEO satellite network on the operation of ground user equipment, it is essential to choose locations separated by considerable distances. Therefore, in this paper, the simulations of the integrated satellite-terrestrial network are run on a scenario that includes the source city located in Sydney, Australia, the destination city located in Cape Town, South Africa, and the 30-satellite constellation distributing in 6 orbital planes. Each orbital plane is equipped with 5 satellites that are distributed uniformly as shown in Fig. 2. Particularly, the satellites information are generated by the Systems Tool Kit (STK). The altitude and inclination of each satellite are 880 km and 45 degree, respectively. The average transmission power of each user equipment, TST, and LEO satellite is 1.25 W, 2 W, and 5 W, separately. The average size of a request is 1 MB. The bandwidth of 5 GHz-band communications is 20 MHz, while Ka-band communications have a bandwidth of 250 MHz. Besides, the center carrier frequency of the Ka-band is considered as 30 GHz. The noise density for terrestrial and satellite-terrestrial communications is 174 dB/Hz. The gain of the satellite antenna is 43.3 dBi. The duration of each time slot is 2 seconds, and the total observation period is 1200 seconds. The DRO framework incorporates 50 samples. Besides, $\alpha = 3$ in the fifteenth constraint in (27). The system parameters are summarized in Table I.

B. Experimental Results

Fig. 3 depicts the objective total latency under the proposed DRO model along with the ambiguity set size. The actual radius of the offloading and downloading ambiguity set is configured as the minimum delay of the uplink path and downlink path via TST embedded in vehicles by multiplying the amplification factor θ shown on the x-axis of Fig. 3. To be specific, the factor θ controls how large the ambiguity set and adjusts the range of the uncertain latency via building and vehicle routing path. Here, we assume $K = 4$ user equipment and 5 tasks to be delivered by each user equipment. We can observe that as the size of the Wasserstein ambiguity set increases, the amount of uncertainty in predicting vehicle movement increases, resulting in an increase in latency. The size of an ambiguity set can significantly differ based on the particular circumstances and the type of uncertainty present.

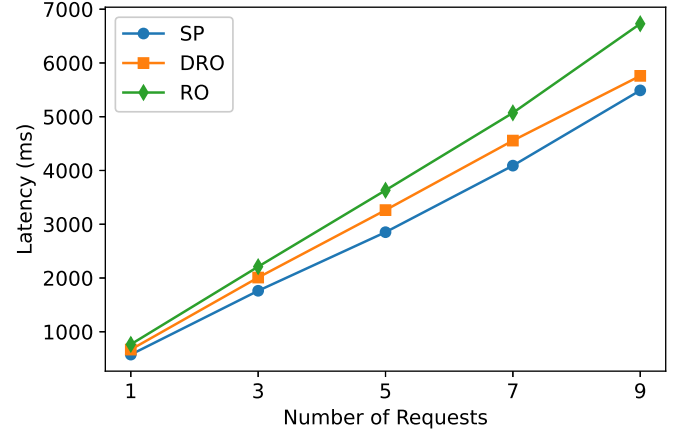


Fig. 4: Latency vs. number of requests. Each line shows the total latency for routing the different number of requests generated by four source user equipment to the corresponding destination user equipment.

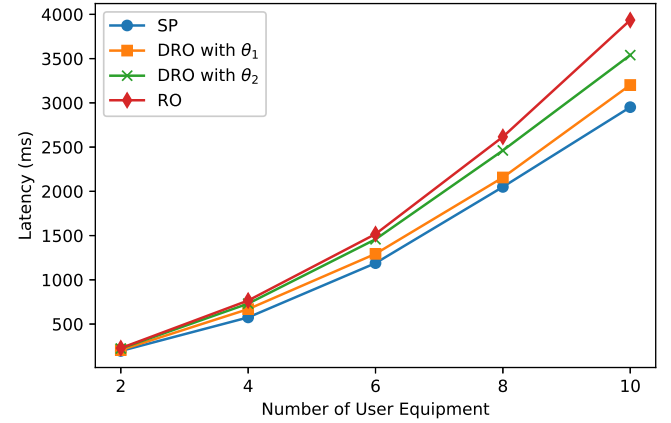


Fig. 5: Latency vs. number of user equipment. Each line shows the total latency for routing one task generated by different number of source user equipment to the corresponding destination user equipment.

In this paper, we utilize the different sizes of the latency ambiguity set to represent varying degrees of uncertainty in vehicle movement directions. By adjusting the ambiguity set, we can evaluate the influence of uncertainty on prediction latency via vehicle path.

Fig.4 illustrates the total required latency for delivering requests with varying volumes generated by four user equipment. Since SP uses the true latency probability distribution of the uplink and downlink path involving automobiles, it has the lowest latency of the three approaches. For the RO scheme, only the range of the uncertain offloading and downloading latency through vehicles is known, resulting in the largest delay. Moreover, DRO is an extension of both SP and RO. Thus, the resulting latency is situated between SP and RO. SP focuses on the average result, potentially overlooking extreme

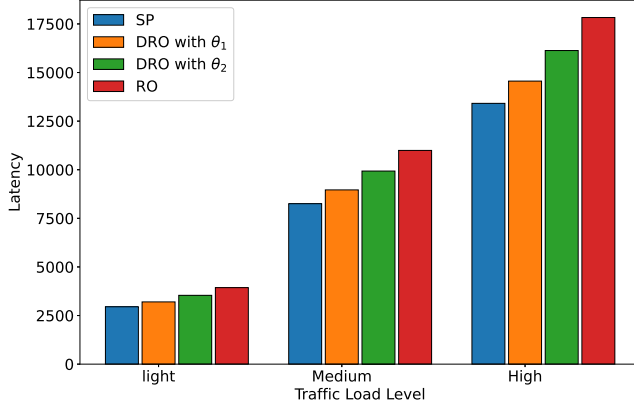


Fig. 6: Latency vs. traffic load. Each bar shows the total latency for routing the one task generated by the 10 source user equipment to the corresponding destination user equipment, but with different information sizes.

or unforeseen conditions. RO is designed to guard against the worst possible scenarios within a predefined uncertainty set, usually leading to over-conservative decision-making. DRO provides a balanced approach between both and is a trade-off in decision-making under uncertainty. It is more robust than SP, including extreme and unforeseen conditions, and less conservative than RO, avoiding over-conservative decision-making. This balance makes DRO an effective strategy for managing the uncertainties during offloading and downloading routing in our integrated satellite-terrestrial networks. In addition, the required routing delay increases as the number of forwarded requests increases.

In Fig. 5, we present the network latency for different user equipment with the same number of tasks in the source city. As more user equipment transmit tasks, additional time is required to complete the assignment. The proposed DRO method possesses the advantages of the SP and RO method, which finds the worst-case expected latency under an uncertain probability distribution with unknown random variables. In addition, we compare the routing response time under two distinct ambiguity set sizes. We set two different amplification factors θ_1 and θ_2 to represent the degree of uncertainty in the proposed DRO model, where $\theta_2 > \theta_1$. Since the reduced size of the ambiguity set brings the possible latency probability distributions closer to or even equal to the known latency probability distribution, the delay for the DRO method with the amplification factor θ_1 will be comparable to that of the SP scheme. Alternately, as the size of the ambiguity set increases, it contains a greater number of potential probability distributions. When the size is sufficiently large, there appears to be no trend in the probability distribution. In this case, the latency outcome for transmitting tasks from source user equipment to respective target user equipment approaches the RO mechanism.

In addition, Fig. 6 illustrates the total routing latency of each method under light, medium, and heavy traffic load conditions.

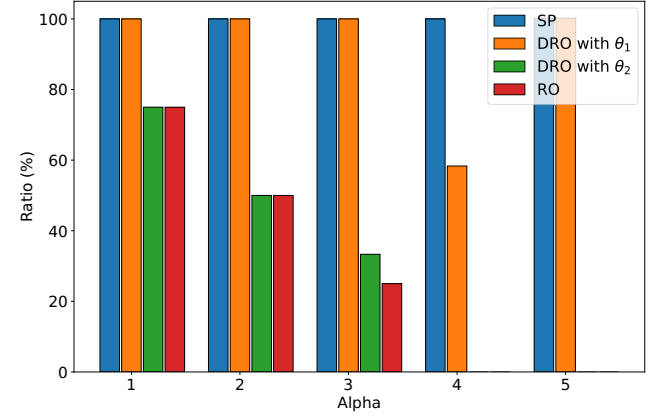


Fig. 7: Ratio vs. Alpha. Each bar shows the total latency for routing three tasks generated by four source user equipment to the corresponding destination user equipment.

For medium and heavy traffic loads, the average data size is 3 MB and 5 MB, correspondingly. The larger the delivered requests, the greater the required latency. Again, DRO performance falls between that of SP and RO. The ambiguity set, controlled by parameter θ_1 , is considered smaller than the one controlled by parameter θ_2 . Even with medium and high traffic loads, the resultant latency for θ_1 and θ_2 is close to SP and RO, separately.

To further discuss the effect of α , which indicates the maximum number of tasks routed through a building during each time slot, we carry out a simulation to figure out the proportion of tasks delivered through uncertain delay paths via vehicles in the case of four user equipment, each with three tasks, as shown in Fig. 7. Due to the fact that the SP method is aware of the true uncertain latency probability distribution, all tasks are transmitted through automobiles with the smallest possible delay. From a different perspective, SP can also be regarded as the DRO method with an extremely small radius of the ambiguity set. Assume that the size amplification factor $\theta_1 < \theta_2$. Except for $\alpha = 4$, all tasks are routed through vehicles for the DRO approach with θ_1 due to the smaller ambiguity set size. The routing decision is affected not only by the scenario setting, i.e. setting of α , but also by the ambiguity set size itself. When the size of the Wasserstein ambiguity set increases, there is a greater likelihood that the uncertain routing latency via vehicles will increase. As a consequence, the model prefers to select the direct uplink path via buildings. As α increases, it is apparent that the ratio of the DRO method with θ_2 and the RO approach decreases. Furthermore, when $\alpha = 4$ and 5, both methods determine to deliver requests via direct paths.

In Fig. 8, we extend our simulation model from a single source-destination city pair (Sydney, Australia to Cape Town, South Africa) to include two additional pairs: Tokyo, Japan to Cape Town, South Africa, and Tokyo, Japan to Sydney, Australia. The user equipment is distributed across the three source cities, and requests are required to be transmitted to

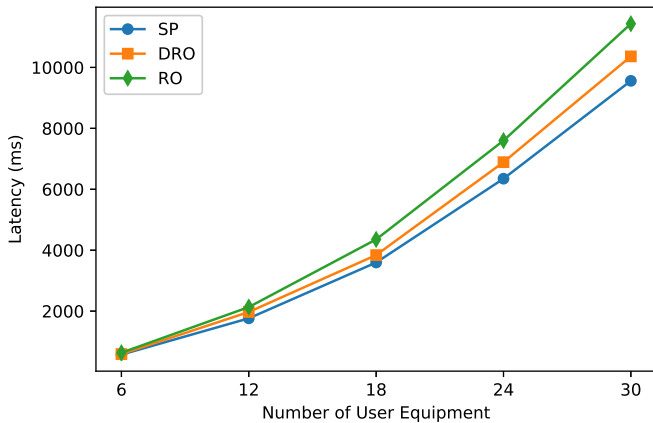


Fig. 8: Latency vs. number of user equipment. Each line shows the total latency for routing one task generated by different number of source user equipment to the corresponding destination user equipment. The three source-destination pairs include from Sydney, Australia to Cape Town, South Africa, from Tokyo, Japan to Cape Town, South Africa and from Tokyo, Japan to Sydney, Australia.

the corresponding target user equipment in the destination cities. By conducting the simulation, we demonstrate that our proposed DRO model can route requests to user equipment located in multiple cities. Again, the simulation findings indicate that the DRO model estimates a total routing latency that falls between the values predicted by the SP and RO models. This highlights the effectiveness of DRO in balancing the trade-offs between risk and performance. DRO provides a well-rounded solution by reducing the inflexibility of RO while simultaneously tackling the shortcomings associated with SP's distributional assumptions.

V. CONCLUSIONS

In this paper, we proposed a DRO-integrated satellite-terrestrial network routing model for delivering requests from the source to the destination user equipment via the LEO satellite constellation with the shortest possible worst-case latency. Due to the unpredictable vehicle movement, there is an intermittent transmission connection between user equipment, automobiles, and satellites during task offloading, and downloading. As a result, the offloading and downloading latency associated with automobiles are uncertain, which can be represented by Wasserstein ambiguity sets. Since the original DRO formulation is difficult to solve directly, we use the duality property to rewrite the inner expectation problem as a finite convex problem in order to derive the tractable optimization formulation. The numerical results demonstrate that the proposed DRO method can route tasks in satellite-terrestrial networks. In addition, the estimated latency varies as the size of the ambiguity set is adjusted. The proposed strategy outperforms the conservative RO approach but performs comparably to the SP strategy with actual true probability distribution.

ACKNOWLEDGEMENT

This work is partially supported by NSF CNS-2107216, CNS-2128368, CMMI-2222810, ECCS-2302469, US Department of Transportation, Toyota, Amazon and Japan Science and Technology Agency (JST) Adopting Sustainable Partnerships for Innovative Research Ecosystem (ASPIRE) JP-MJAP2326. Also, this work is partially supported by grant #80NSSC22K0259 from the National Aeronautics and Space Administration.

REFERENCES

- [1] L. Kuang, Z. Feng, Y. Qian, and G. Giambene, "Integrated terrestrial-satellite networks: Part two," *China Communications*, vol. 15, no. 8, pp. 4–6, Aug. 2018.
- [2] Y. Henri, "The OneWeb satellite system," *Handbook of Small Satellites: Technology, Design, Manufacture, Applications, Economics and Regulation*, pp. 1–10, Feb. 2020.
- [3] M. Handley, "Delay is not an option: Low latency routing in space," in *Proceedings of the 17th ACM Workshop on Hot Topics in Networks*, ser. HotNets, New York, NY, Nov. 2018, p. 85–91.
- [4] —, "Using ground relays for low-latency wide-area routing in mega-constellations," in *Proceedings of the 18th ACM Workshop on Hot Topics in Networks*, ser. HotNets, New York, NY, Nov. 2019, p. 125–132.
- [5] C. Qiu, H. Yao, F. R. Yu, F. Xu, and C. Zhao, "Deep Q-learning aided networking, caching, and computing resources allocation in software-defined satellite-terrestrial networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5871–5883, Mar. 2019.
- [6] K. Lin, D. Wang, L. Hu, M. S. Hossain, and G. Muhammad, "Virtualized QoS-driven spectrum allocation in space-terrestrial integrated networks," *IEEE Network*, vol. 33, no. 1, pp. 58–63, Jan. 2019.
- [7] H. A. Shah, L. Zhao, and I.-M. Kim, "Joint network control and resource allocation for space-terrestrial integrated network through hierarchical deep actor-critic reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 4943–4954, May 2021.
- [8] T. Van Chien, E. Lagunas, T. M. Hoang, S. Chatzinotas, B. Ottersten, and L. Hanzo, "Power allocation for space-terrestrial cooperation systems with statistical CSI," in *2022 IEEE Global Communications Conference*, Dec. 2022, pp. 3284–3289.
- [9] Y. Yin, C. Huang, N. N. Xiong, D.-F. Wu, and S. Huang, "Joint dynamic routing and resource allocation in satellite-terrestrial integrated networks," *Computer Networks*, vol. 231, p. 109823, May 2023.
- [10] W. Abderrahim, O. Amin, M.-S. Alouini, and B. Shihada, "Latency-aware offloading in integrated satellite terrestrial networks," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 490–500, Apr. 2020.
- [11] J. Li, K. Xue, D. S. L. Wei, J. Liu, and Y. Zhang, "Energy efficiency and traffic offloading optimization in integrated satellite/terrestrial radio access networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2367–2381, Jan. 2020.
- [12] D. Han, Q. Ye, H. Peng, W. Wu, H. Wu, W. Liao, and X. Shen, "Two-timescale learning-based task offloading for remote IoT in integrated satellite-terrestrial networks," *IEEE Internet of Things Journal*, pp. 1–1, Jun. 2023.
- [13] Z. Zhang, W. Zhang, and F.-H. Tseng, "Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques," *IEEE Network*, vol. 33, no. 1, pp. 70–76, Jan. 2019.
- [14] R. Xie, Q. Tang, Q. Wang, X. Liu, F. R. Yu, and T. Huang, "Satellite-terrestrial integrated edge computing networks: Architecture, challenges, and open issues," *IEEE Network*, vol. 34, no. 3, pp. 224–231, Mar. 2020.
- [15] T. Chen, J. Liu, Q. Ye, W. Zhuang, W. Zhang, T. Huang, and Y. Liu, "Learning-based computation offloading for iort through ka/q-band satellite-terrestrial integrated networks," *IEEE Internet of Things Journal*, vol. 9, no. 14, pp. 12 056–12 070, Jul. 2022.
- [16] N. Torkzaban, A. Gholami, J. S. Baras, and C. Papagianni, "Joint satellite gateway placement and routing for integrated satellite-terrestrial networks," in *IEEE International Conference on Communications (ICC)*, Dublin, Ireland, Jul. 2020.
- [17] D. Zhou, M. Sheng, J. Wu, J. Li, and Z. Han, "Gateway placement in integrated satellite-terrestrial networks: Supporting communications and internet of remote things," *IEEE Internet of Things Journal*, vol. 9, no. 6, pp. 4421–4434, Mar. 2022.

- [18] Y. Xiao, T. Zhang, and L. Liu, "Addressing subnet division based on geographical information for satellite-ground integrated network," *IEEE Access*, vol. 6, pp. 75 824–75 833, Nov. 2018.
- [19] Q. Guo, R. Gu, T. Dong, J. Yin, Z. Liu, L. Bai, and Y. Ji, "SDN-based end-to-end fragment-aware routing for elastic data flows in LEO satellite-terrestrial network," *IEEE Access*, vol. 7, pp. 396–410, Jan. 2019.
- [20] H. Xu, D. Li, M. Liu, G. Han, W. Huang, and C. Xu, "A hybrid routing algorithm in terrestrial-satellite integrated network," in *IEEE/CIC International Conference on Communications in China (ICCC)*, Chongqing, China, Aug. 2020, pp. 90–95.
- [21] K. Guo, D. Wang, H. Zhi, Y. Lu, and Z. Jiao, "A trusted resource-based routing algorithm with entropy estimation in integrated space-terrestrial network," *IEEE Access*, vol. 8, pp. 122 456–122 468, Jul. 2020.
- [22] F. Wang, D. Jiang, Z. Wang, Z. Lv, and S. Mumtaz, "Fuzzy-CNN based multi-task routing for integrated satellite-terrestrial networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 1913–1926, Dec. 2022.
- [23] J. R. Birge and F. Louveaux, *Introduction to stochastic programming*. Springer Science & Business Media, 2011.
- [24] L. A. Hannah, "Stochastic optimization," *International Encyclopedia of the Social & Behavioral Sciences*, vol. 2, pp. 473–481, 2015.
- [25] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust optimization*. Princeton university press, 2009, vol. 28.
- [26] C. Zhao and Y. Guan, "Data-driven risk-averse stochastic optimization with wasserstein metric," *Operations Research Letters*, vol. 46, no. 2, pp. 262–267, March 2018.
- [27] H. Nakao, R. Jiang, and S. Shen, "Distributionally robust partially observable markov decision process with moment-based ambiguity," *SIAM Journal on Optimization*, vol. 31, no. 1, pp. 461–488, May 2021.
- [28] P. Xiong, P. Jirutitijaroen, and C. Singh, "A distributionally robust optimization model for unit commitment considering uncertain wind power generation," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 39–49, Jan. 2017.
- [29] S. Wang, C. Zhao, L. Fan, and R. Bo, "Distributionally robust unit commitment with flexible generation resources considering renewable energy uncertainty," *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4179–4190, February 2022.
- [30] Y. Zhang, S. Shen, and S. A. Erdogan, "Distributionally robust appointment scheduling with moment-based ambiguity set," *Operations Research Letters*, vol. 45, no. 2, pp. 139–144, Mar. 2017.
- [31] C. Shang and F. You, "Distributionally robust optimization for planning and scheduling under uncertainty," *Computers & Chemical Engineering*, vol. 110, pp. 53–68, Feb. 2018.
- [32] Z. Ling, F. Hu, H. Zhang, and Z. Han, "Age-of-information minimization in healthcare iot using distributionally robust optimization," *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 16 154–16 167, Sep. 2022.
- [33] D. Zhou, M. Sheng, B. Li, J. Li, and Z. Han, "Distributionally robust planning for data delivery in distributed satellite cluster network," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3642–3657, May 2019.
- [34] Y. Liu, H. Zhang, B. Di, J. Wu, and Z. Han, "Deployment for high altitude platform systems with perturbation: Distributionally robust optimization approach," *IEEE Communications Letters*, vol. 26, no. 5, pp. 1126–1130, Mar. 2022.
- [35] J. Zhao, L. Yu, K. Cai, Y. Zhu, and Z. Han, "RIS-aided ground-aerial noma communications: A distributionally robust drl approach," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1287–1301, Jan. 2022.
- [36] Y. Chen, B. Ai, Y. Niu, H. Zhang, and Z. Han, "Energy-constrained computation offloading in space-air-ground integrated networks using distributionally robust optimization," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 12 113–12 125, Sept. 2021.
- [37] Z. Ling, F. Hu, Y. Zhang, L. Fan, F. Gao, and Z. Han, "Distributionally robust chance-constrained backscatter communication-assisted computation offloading in wbans," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3395–3408, Feb. 2021.
- [38] Y. Zi, L. Fan, X. Wu, J. Chen, and Z. Han, "Distributionally robust optimal sensor placement method for site-scale methane-emission monitoring," *IEEE Sensors Journal*, vol. 22, no. 23, pp. 23 403–23 412, Dec. 2022.
- [39] L. Wu and M. Hifi, "Discrete scenario-based optimization for the robust vehicle routing problem: The case of time windows under delay uncertainty," *Computers & Industrial Engineering*, vol. 145, p. 106491, 2020.
- [40] X. Wang, Y. Weng, and H. Gao, "A low-latency and energy-efficient multimetric routing protocol based on network connectivity in vanet communication," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 4, pp. 1761–1776, 2021.
- [41] J. Lu, X. Wang, L. Zhang, and X. Zhao, "Fuzzy random multi-objective optimization based routing for wireless sensor networks," *Soft Computing*, vol. 18, pp. 981–994, 2014.
- [42] K.-C. Tsai, L. Fan, R. Lent, L.-C. Wang, and Z. Han, "Integrated satellite-terrestrial routing using distributionally robust optimization," in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 277–282.
- [43] S. Tani, K. Motoyoshi, H. Sano, A. Okamura, H. Nishiyama, and N. Kato, "An adaptive beam control technique for diversity gain maximization in LEO satellite to ground transmissions," in *IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, May 2016.
- [44] P. Series, "Propagation data and prediction methods required for the design of earth-space telecommunication systems," *Recommendation ITU-R*, pp. 618–12, Jul. 2015.
- [45] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations research*, vol. 58, no. 3, pp. 595–612, Jan. 2010.
- [46] W. Wiesemann, D. Kuhn, and M. Sim, "Distributionally robust convex optimization," *Operations Research*, vol. 62, no. 6, pp. 1358–1376, Oct. 2014.
- [47] A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen, "Robust solutions of optimization problems affected by uncertain probabilities," *Management Science*, vol. 59, no. 2, pp. 341–357, Feb. 2013.
- [48] R. Jiang and Y. Guan, "Data-driven chance constrained stochastic program," *Mathematical Programming*, vol. 158, no. 1–2, pp. 291–327, 2016.
- [49] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, no. 1–2, pp. 115–166, Sep. 2018.
- [50] R. Gao and A. Kleywegt, "Distributionally robust stochastic optimization with Wasserstein distance," *Mathematics of Operations Research*, Aug. 2022.
- [51] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951. [Online]. Available: <http://www.jstor.org/stable/2236703>
- [52] M. Stemm and R. H. Katz, "Vertical handoffs in wireless overlay networks," *Mobile Networks and Applications*, vol. 3, pp. 335–350, 1998.
- [53] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, no. 1, pp. 115–166, Sep. 2018.



Kai-Chu Tsai received her B.S. degree and M.S. degree in communication engineering, from National Central University and National Taiwan University in 2015 and 2017, respectively. She is currently working toward the Ph.D. degree in the Electrical and Computer Engineering Department, University of Houston, USA. In 2021, she visited the Department of Electrical and Computer Engineering, National Yang Ming Chiao Tung University, Taiwan, doing research on network routing and optimization. Her research interests include telecommunication, reinforcement learning and optimization.



Lei Fan (Senior Member, IEEE) is an Assistant Professor in the Department of Engineering Technology and also holds a joint appointment with the Department of Electrical and Computer Engineering at the University of Houston. Prior to this position, he worked in the power industry for several years. He earned his Ph.D. in operations research from the Department of Industrial and Systems Engineering at the University of Florida. His research interests encompass power system operations and planning, electricity markets, optimization algorithms, complex network systems, and quantum computing.



Ricardo Lent (M'92, SM'14) received his M.Sc. and Ph.D. degrees in computer science from the University of Central Florida, USA, in 2003, his M.Sc. in Telecommunications Engineering from *Universidad Nacional de Ingeniería* and his B.Sc. in Electronic Engineering from Ricardo Palma University. He currently holds the position of Associate Professor in the Department of Engineering Technology, College of Engineering at the University of Houston. Prior to this role, he served as a Research Fellow in the Department of Electrical and Electronic Engineering

at Imperial College London, as well as in related industry roles. As the recipient of a NASA Early Career Faculty (ECF) and a NASA Fellowship, he directs his research efforts towards advancing space communications, delay-tolerant networking, cognitive networking, and green computing.



Li-Chun Wang (M'96-SM'06-F'11) received Ph.D. degree from the Georgia Institute of Technology in 1996. From 1996 to 2000, he served as a Senior Researcher at the Wireless Communications Research Institute at AT&T Labs. He is currently the Dean of the College of Electrical Engineering and a Distinguished Professor in the Department of Electrical Engineering at National Yang Ming Chiao Tung University. Dr. Wang was elected as a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) in 2011 for his contributions to the design of cellular architectures

and wireless resource management in wireless networks. Dr. Wang has received numerous awards and honors, including the Distinguished Research Awards from National Science and Technology Council twice (2012 and 2016), the Future Tech Award from the National Science and Technology Council (2021), the Chinese Institute of Engineers (CIE) Outstanding Electrical Engineering Professor Award (2022), the Outstanding Engineering Professor Award from the Chinese Institute of Electrical Engineering (2009), the K.T. Li Fellow Award (2021) and the Medal of Honor (2024) from the Institute of Information & Computing Machinery (iICM), the Outstanding ICT. Elite Award (2020), the Y. Z. Hsu Scientific Paper Award (2013), and the Y. Z. Hsu Scientific Chair Professor (2023). Dr. Wang has made significant contributions to the research fields of wireless communication and information technology. He has served as an IEEE Tutorial Speaker multiple times, promoting international cooperation and talent cultivation. According to Google Scholar, Dr. Wang's research works have been cited over 9,800 times with an h-index of 49. He was listed in the "2020 Annual Global Top 2% Scientists" and "Lifetime Scientific Impact Rankings" by Stanford University. He is also ranked as a top Taiwanese international scholar in the field of computer science by the Guide2 Research website. Dr. Wang serves as the Director of the Chunghwa Telecom-NYCU Innovation Research Center and the NYCU-IBM IoT Research Center. He has collaborated with numerous domestic and international companies and holds 49 domestic and international patents, sixteen of which have been applied in commercial products. He is currently an Associate Editor of the IEEE Transactions on Wireless Communications and the IEEE Transactions on Cognitive and Communication Networking. His recent research interests lie in data-driven intelligent wireless communications, brain technology, and sustainable development. He has published over 300 journal and conference papers and co-edited the book "Key Technologies for 5G Wireless Communications" (Cambridge University Press, 2017).



Zhu Han (S'01-M'04-SM'09-F'14) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2003, respectively.

From 2000 to 2002, he was an R&D Engineer of JDSU, Germantown, Maryland. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an assistant professor at Boise State University, Idaho. Currently, he is a John and Rebecca Moores Professor in the Electrical and Computer Engineering Department as well as in the Computer Science Department at the University of Houston, Texas. Dr. Han's main research targets on the novel game-theory related concepts critical to enabling efficient and distributive use of wireless networks with limited resources. His other research interests include wireless resource allocation and management, wireless communications and networking, quantum computing, data science, smart grid, carbon neutralization, security and privacy. Dr. Han received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the Journal on Advances in Signal Processing in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (best paper award in IEEE JSAC) in 2016, IEEE Vehicular Technology Society 2022 Best Land Transportation Paper Award, and several best paper awards in IEEE conferences. Dr. Han was an IEEE Communications Society Distinguished Lecturer from 2015 to 2018 and ACM Distinguished Speaker from 2022 to 2025, AAAS fellow since 2019, and ACM Fellow since 2024. Dr. Han is a 1% highly cited researcher since 2017 according to Web of Science. Dr. Han is also the winner of the 2021 IEEE Kiyo Tomiyasu Award (an IEEE Field Award), for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks."