Incentive Systems for Fleets of New Mobility Services

Ali Ghafelebashi, Meisam Razaviyayn, and Maged Dessouky

***Abstract*—Traffic congestion has become an inevitable challenge in large cities due to population increases and the expansion of urban areas. Various approaches are introduced to mitigate traffic issues, encompassing from expanding the road infrastructure to employing demand management. Congestion pricing and incentive schemes are extensively studied for traffic control in traditional networks where each driver/rider is a network "player". In this setup, drivers'/riders' "selfish" behavior hinders the network from reaching a socially optimal state. In future mobility services, on the other hand, a large portion of drivers/vehicles may be controlled by a small number of companies/organizations. In such a system, offering incentives to organizations can potentially be much more effective in reducing traffic congestion rather than offering incentives directly to drivers. This paper studies the problem of offering incentives to organizations to change the behavior of their individual drivers (or individuals relying on the organization's services). We developed a model where incentives are offered to each organization based on their aggregated travel time loss across all drivers/riders in that organization. Such an incentive offering mechanism requires solving a large-scale optimization problem to minimize the system-level travel time. We propose an efficient algorithm for solving this optimization problem. Numerous experiments on Los Angeles County traffic data reveal the ability of our method to reduce system-level travel time by up to 7.15%. Moreover, our experiments show that incentivizing organizations can be up to 7 times more cost-effective than incentivizing individual drivers when aiming for maximum travel time reduction.**

***Index Terms*—New Mobility Services, Congestion Reduction, Incentivizing Organizations, Travel Demand Management.**

I. Introduction

TODAY, traffic congestion is one of the major issues in metropolitan areas across the globe. Traffic congestion declines the overall quality of life, leads to significant economic losses, degrades air quality, and escalates health vulnerabilities due to emissions [1–4]. This paper devises a novel mechanism around incentives. The core objective of this mechanism is to change the behavioral patterns of individual drivers within organizations by incentivizing organizations.

Incentive-based congestion reduction methodologies overlap with pricing methods, including taxes and fees for road access [5–14]. These strategies encourage individuals to avoid congested routes, reducing traffic buildup. Various determinants underpin the design of these pricing frameworks, encompassing temporal aspects [15], spatial metrics [16], and vehicular attributes [17, 18]. Although promising, market-oriented pricing and taxation face challenges due to equity concerns, policy complexity, and implementation uncertainties [19–25].

Another approach within the area of pricing mechanisms involves the adoption of tradable credits (TCs) or tradable mobility permits (TPMs) [26–29]. [30] provides a theoretical analysis of the benefits of tradable credits. This methodology has been implemented within some economic sectors, exemplified by its use in the airport slot market [31]. Nevertheless, implementing these cap-and-trade programs in personal travel and daily commutes is hindered by design challenges [32, 33].

Recently, there has been a heightened focus on incentivization strategies. Compared to fee-based methods, reward-based policies can be more popular [34]. Moreover, the efficacy of incentivizing positive actions over punishing negative ones is evidenced in the psychological concept of reactance [34]. While rewarding policies have proven effective in altering individual behavior [35, 36], the transportation sector has underexplored these incentives.

There have been several studies that explored the use of incentivization to reduce traffic congestion, such as the INSTANT project [37], the CAPRI project of Stanford [38], series of studies in the Netherlands [39], and the "Metropia" platform [40]. In a recent study, [41] showed the effectiveness of ridesharing incentivization in congestion reduction. Incentivizing off-peak hour driving is examined via public and private central planners (policymakers) in [42]. However, congestion reduction by offering incentives to organizations has not been studied by any of the previous studies. Although initial success is shown in reward-based strategies, enduring behavior change is not always guaranteed [43].

In traditional congestion pricing and incentive offering mechanisms, incentives are offered directly to individual drivers to influence their decisions, such as departure time and routing (Fig. 1 (a)). In mobility services, many of these decisions may be directly (or indirectly) made by organizations providing

Manuscript received December 7, 2023. This study was funded by a grant from the National Center for Sustainable Transportation (NCST), supported by the U.S. Department of Transportation's University Transportation Centers Program. The contents of this project reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The U.S. Government and the State of California assume no liability for the contents or use thereof. Nor does the content necessarily reflect the official views or policies of the U.S. Government and the State of California. This paper does not constitute a standard, specification, or regulation. This paper does not constitute an endorsement by the California Department of Transportation (Caltrans) of any product described herein.

Ali Ghafelebashi, Meisam Razaviyayn, and Maged Dessouky are with the Daniel J. Epstein Department of Industrial & Systems Engineering, University of Southern California, 3715 McClintock Avenue, Los Angeles, CA 90089, United States.

Email addresses: ghafeleb@usc.edu (Ali Ghafelebashi), razaviya@usc.edu (Meisam Razaviyayn), maged@usc.edu (Maged Dessouky).

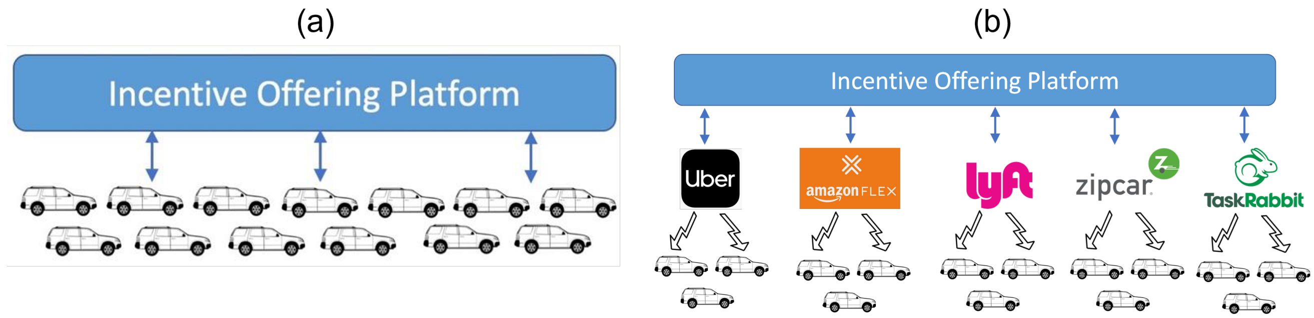

Fig. 1. (a) Traditional platforms for offering incentives: incentives are offered to individual drivers in the system. (b) Presented platform for offering incentives: incentives are offered to new mobility services to change their drivers' behavior.

different transportation services. For example, navigation apps, which are regularly used by almost 70% of smartphone users [44, 45], influence the routing decision of millions of drivers daily. Another example is crowdsourcing delivery platforms such as Amazon Flex, Instacart, and Doordash. According to a recent study [46], the revenue of DoorDash during the fourth quarter (Q4) of 2022 increased by 40% to $1.8 billion from $1.3 billion in revenue that it recorded during Q4 2021. Another example of such organizations is ride-hailing organizations such as Uber and Lyft. According to a report by Uber for the fourth quarter of 2022 [47], the number of gross bookings increased from 12% in Q4 2021 to 17.7% in Q4 2022. Today, many routing decisions are made by individual drivers. With the future emergence of autonomous vehicles, it is possible that organizations may now own the fleet of vehicles and control their routing.

Intuitively, since organizations have more flexibility and more power to change the traffic, incentivizing organizations is expected to be more efficient than incentivizing individual drivers. Furthermore, an organization has more options in balancing the route selection across the large pool of drivers employed by the organization. Motivated by this idea, this paper develops an incentive offering mechanism to organizations to indirectly (or directly) influence the behavior of individual drivers (Fig. 1 (b)). In a different study, [48] utilizes the traditional incentive offering framework (Fig. 1 (a)) to provide an algorithm to offer personalized incentives to drivers to reduce traffic congestion by changing the routing decision of the drivers. These incentives could be personalized based on user preferences. Both individual level and organization level incentivization mechanisms will not charge participant drivers who gain (by arriving early) because, in this paper, we only focus on a reward system instead of a cost-based system.

In contrast to the individual-level incentivization presented in [37–39, 48], our incentivization framework for organizations addresses a broader spectrum of challenges and complexities:

1) Numerical experiments in [48] show that the monetary benefit from reduced travel time based on the Value of Time (VOT) exceeds the incentivization cost. However, their model does not depend on VOT. Our model uses VOT to compute the monetary value of organizations' time loss and compensate for it through incentivization.
2) Some incentivization studies [37–39] offered static rewards based on fixed rules for all the participants. However, our model utilizes different VOTs to compute the incentive offer for different organizations. Note that different organizations can have different VOTs due to the unique nature of their service. Hence, our model offers incentives to organizations such that they can compensate the organizations' time loss based on their VOT.
3) [48] offers personalized incentives, but they are selected from a discrete set of incentive choices that are fixed before solving the problem. However, our model employs a continuous variable to calculate the value of the required incentive. This variable depends on VOT and the amount of time lost by drivers. As we are not limited to a discrete set of incentives, our incentivization cost can be more cost-efficient. Note that the variability in incentive values introduces more complexity to our optimization problem because of the larger variable size.
4) [48] does not consider fairness and time delivery constraints (fair assignment of drivers to slower and faster routes when they share the same origin and destination simultaneously). In contrast, our model addresses these limitations by preventing the diversion of drivers to routes with significant time disparities, thus ensuring a route assignment based on fairness and time delivery constraints.

Our framework will follow this three-step procedure:
Step 1) The central planner receives organizations' demand estimates for the next time interval (e.g., the next few hours).
Step 2) The central planner incentivizes organizations to change their routes and travel time.
Step 3) Observe organizations' response and go back to Step 1 for the next time interval.
The central planner (which is referred to as "Incentive Offering Platform" in Fig. 1 (b)) continually repeats this three-step process in the network for every time interval. A detailed process description is provided in Fig. 2.

Although an incentive system targets the reward to the organization, the proposed approach can be used by a wide variety of transportation organizations, including those that have drivers as employees, use autonomous vehicles, or use gig workers as drivers. Furthermore, the transport organization can be a freight delivery company. In delivery organizations, there are no passengers that need to be incentivized. In autonomous vehicle organizations, there are no drivers that need to be

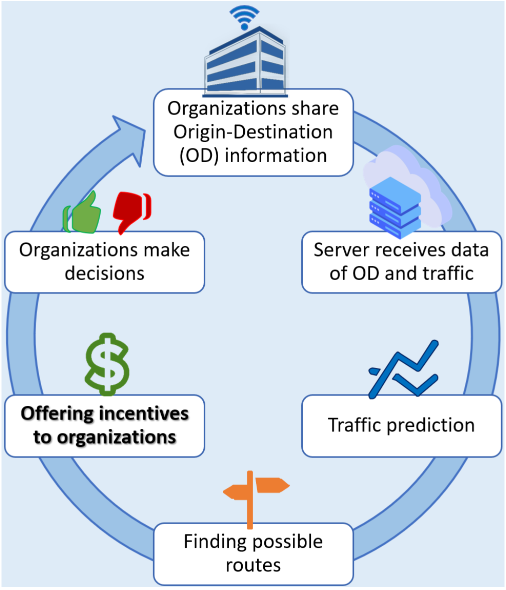

Fig. 2. Detailed description of the incentivization process.

incentivized. In ride-hailing organizations, it encompasses both drivers and passengers. In the categories where gig workers or passengers participate, they are not obligated to accept the ride with a longer travel time. Hence, they can accept the incentivized longer road or reject it. Participant drivers/passengers who reject the incentivized longer route can take the shortest route. As previously mentioned, the platform is designed to provide payments to organizations. In cases where the organization employs gig workers or involves passengers, the organization will have to pass on some of these incentives to the gig workers and passengers to incentivize them to take longer routes. Overall, the incentivization platform aims to improve system efficiency by moving toward System Optimal (SO) by changing organizations' routes via incentives.

The rest of this paper is structured as follows. Section II motivates the advantage of incentivizing organizations compared to incentivizing individuals. Section III introduces the basic notations and describes our incentive offering mechanism for congestion reduction. We formulate an optimization problem to find the "optimal" incentive offering strategy. We then propose an algorithm for solving this optimization problem efficiently in Section IV. Numerical experiment results for the model using Los Angeles County data are detailed in Section V. Concluding remarks are discussed in Section VI. The scalability of the presented platform is discussed in Appendix E

II. Why Offering Incentives to Organizations Rather Than Individuals?

Our methodology is incentivizing organizations (rather than individual drivers). Let us first motivate the benefit of this strategy via a simple example. Consider the subnetwork $\tilde{\mathcal{G}}$ at Fig. 3 as a subset of a larger network. Links a and b are routes between Origin-Destination (OD) nodes v_1 (origin) and v_2 (destination). The travel times of a and b are 25 and 30 minutes at User Equilibrium (UE), respectively. Assume 20

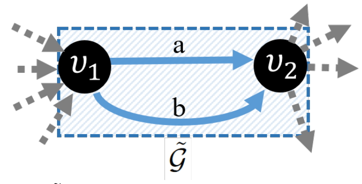

Fig. 3. Subnetwork $\tilde{\mathcal{G}}$ (selected in blue dashed rectangle).

drivers start traveling from v_1 to v_2 at the same time. If travel time is the most important factor in their utility, they will select v_1 because it is the fastest route at UE. Assume we have found the System Optimal (SO) strategy for the entire network, and we need 15 out of the 20 drivers to select b instead of a to achieve SO. At SO, the travel time of route a decreases to 20 minutes (5-minute decrease), and the travel time of route b increases to 35 minutes (5-minute increase). Drivers that use route a save 5 minutes due to a decrease in travel time of route a. Deviated drivers to route b expect to lose 5 minutes because they expect route b to have travel time of 30 minutes. Hence, since we want to deviate 15 drivers to a route with longer travel time (route b in this example), we should compensate for their increased travel time. Assume VOT is \$1/min. Let us compare two scenarios:

(I) All 20 drivers are individual drivers. Since we need to deviate/incentivize 15 drivers and compensate each of them for 5 minutes of their time, we need to spend $\$75 = (5 \text{ min} \times 15) \times \$1/\text{min}$.

(II) All 20 drivers work in the same organization. In this scenario, the organization needs to spend \$75 to alter the decision of the 15 drivers. However, after offering incentives, the travel time of the 5 remaining drivers on route a decreases. Therefore, the organization gains $25 = 5 \times 5$ minutes of time from the drivers who stayed in route a. Overall, the increase and decrease in the drivers' travel times cancel each other out (canceling-out effect). This change only costs the organization 50 minutes of total time. Hence, the compensation cost is $\$50 = 50 \text{ min} \times \$1/\text{min}$ for the organization.

Therefore, we spend 33% less in incentivizing the organization (i.e., scenario (II)) compared to incentivizing the individual drivers (i.e., scenario (I)). This example illustrates that incentivizing organizations can be more cost-effective than incentivizing individual drivers. Note that this observation does not necessarily hold in general games; grouping users in a game does not necessarily lead to a lower-cost Nash equilibrium.

III. Incentive Offering Mechanism and Problem Formulation

Given the origin-destination information of drivers in various organizations, the goal is to find the "optimal" strategy for offering organization-level incentives to them to reduce the traffic congestion of the system. To mathematically state the problem, we begin by defining our notations. A complete list of notations used in this paper can be found in Appendix A. For further details of the notation, an example is provided in Appendix B in the complete version of the work [49].

The traffic network is represented by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Vertices $\mathcal{V}$ of the graph are major ramps and intersections in the network. Vertices are connected by a set of edges $\mathcal{E}$. In our directed graph, the edge direction is determined by the allowable direction of travel on the corresponding road segment, indicating the permissible movement from one node to another for a driver. The adjacency of two nodes is based on the possibility of driving directly from one node to another without visiting any other node. The network comprises a total number of road segments, denoted as $|\mathcal{E}|$, which reflects the cardinality of the set $\mathcal{E}$. A route in the network is a path in the graph and is denoted by a one-hot encoding. In other words, a given route is represented by a vector $\mathbf{r} \in \{0,1\}^{|\mathcal{E}|}$ in which the k-th entry is one if route $\mathbf{r}$ includes the k-th edge and it is zero, otherwise. Let $\mathbf{T} = \{1, \ldots, T\}$ denote the defined time horizon such that $t = 1$ marks the starting time of the system. Traffic volume of road segments at time t is represented by the vector $\mathbf{v}_t \in \mathbb{R}^{|\mathcal{E}|}$ in which the k-th entry is the total number of vehicles of road segment k at time t.

Let $\mathcal{N} = \mathcal{N}_1 \cup \cdots \cup \mathcal{N}_n$ denote the set of all drivers and $\mathcal{N}_i$ denote the set of drivers of organization i. If a driver works for multiple organizations, he or she will be counted as a different driver at each organization. Hence, $\mathcal{N}_1 \cap \cdots \cap \mathcal{N}_n = \emptyset$. For any driver $j \in \mathcal{N}$, let $\mathcal{R}_j \subseteq \{0,1\}^{|\mathcal{E}|}$ denote the set of driver's possible route choices between her/his origin and destination. The binary variable $s_i^{\mathbf{r},j} \in \{0,1\}$ represents the assigned route to the j-th driver of organization i. For this driver and given route $\mathbf{r} \in \mathcal{R}_j$, the variable $s_i^{\mathbf{r},j} = 1$ if route $\mathbf{r}$ is assigned to the j-th driver of organization i; and $s_i^{\mathbf{r},j} = 0$, otherwise. Each driver can only be assigned to one route, i.e., $\sum_{\mathbf{r} \in \mathcal{R}_j} s_i^{\mathbf{r},j} = 1$. Given any routing strategy assigned to drivers, we model the drivers' decision deterministically due to the power of the organizations in enforcing their drivers' routes.

In this paper, we change the routing decision of organizations' drivers by incentivizing their organizations. The incentivization budget can be provided through resources similar to previous studies [37–42] (e.g., government). We assume that organizations will accept our route assignments if the incentive offer can compensate for the change in their total travel time. Notice that when the organizations decide to accept the offer, they have no access to the offered route assignments to the other organizations. Hence, they can only estimate the travel time based on historical data, and they will be compensated based on their loss/gain compared to the historic setting. This compensation is computed by utilizing the VOT at the **organization-level**. For example, in the case of ride-hailing services like Uber and Lyft, the incentivization platform can provide incentives to the organization based on the VOT of drivers and passengers (combined). Next, the organization utilizes the received budget to incentivize passengers (e.g., by reducing the price) and the drivers who accept longer routes (by paying them). Those who reject the incentivized longer route can take the shortest route. Note that an organization's VOT is not necessarily dependent on passengers or drivers because the set of organizations extends beyond the ride-hailing sector. For instance, the VOT for delivery services would be associated with the delivery partners' VOT. Moreover, the VOT of autonomous vehicle organizations like Waymo pertains only to passengers due to the absence of drivers. Our platform includes the flexibility to differentiate VOTs for each organization due to the varied nature of their operations.

In this work, we adopt total travel time as the utility function, while alternative metrics like energy consumption or total carbon emissions can also be considered. We compute the system total travel time by summing the drivers' travel time of all road segments over all time periods in the horizon of interest:

$$F(\hat{\mathbf{v}}) = \sum_{\ell=1}^{|\mathcal{E}|} \sum_{t=1}^{|\mathbf{T}|} \hat{v}_{\ell,t} \theta_{\ell,t}(\hat{v}_{\ell,t}) \tag{1}$$

where $\theta_{\ell,t}$ is the travel time of link ℓ at time t (which itself is a function of the link's traffic volume at that specific time). Here, $\hat{\mathbf{v}}$ is the vector of volume of links in which $\hat{v}_{\ell,t}$ is the $(|\mathcal{E}| \times (t-1) + \ell)^{th}$ element of vector $\hat{\mathbf{v}}$ corresponding to the volume on the ℓ^{th} link at time t. Using the volume vector, we can then calculate the travel time for the links at various times, as outlined below.

Multiple approaches have been proposed to illustrate the relationship between traffic volume and travel time. For instance, the Bureau of Public Roads (BPR) [50] presents a congestion function for road links. This function describes a non-linear connection between the travel time on a road and its traffic volume:

$$\theta(v) = f_{\text{BPR}}(v) = \theta_0 \left(1 + 0.15 \left(\frac{v}{w}\right)^4\right) \tag{2}$$

where $f_{\text{BPR}}(v)$ denotes the travel time for drivers on a road segment based on its traffic volume v; θ_0 represents the segment's free flow travel time; and w is the road segment's practical capacity. In our experiments, to learn the parameters w and θ_0 of the road segments in the Los Angeles area at different times of the day, we utilize the historical traffic data of the road segments. Given the function $\theta(\cdot)$ in (2), to compute the total travel time of the system, one needs to compute the volume at each link. Subsequently, we elucidate the process by which the volume vector is computed within our model.

Volume vector $\hat{\mathbf{v}}$*:* The computation of the volume vector $\hat{\mathbf{v}}$ requires (approximately) estimating the location of the drivers at different times based on their route. By assigning a different route to a driver, the driver's impact on the values of the vector $\hat{\mathbf{v}}$ will be different because the driver's location will change by following a different route. We will begin by introducing our notation for route assignment: Each driver's assigned route is represented by a one-hot encoded vector. Thus, for each driver, we have a binary vector $\mathbf{s}_i^j \in \{0,1\}^{|\mathcal{R}|}$ in which only one element has a value of one, and it corresponds to the assigned route to the j-th driver of organization i. As we need one vector for each driver, we can aggregate all our assignments in a matrix $\mathbf{S} \in \{0,1\}^{|\mathcal{R}| \times |\mathcal{N}|} = [\mathbf{S}_1 \mathbf{S}_2 \ldots \mathbf{S}_n]$ where $\mathbf{S}_i \in \{0,1\}^{|\mathcal{R}| \times |\mathcal{N}_i|}$, which is the assignment matrix of organization i with n being the number of organizations. Elements in a driver's assignment vector that correspond to routes unrelated to their specific origin-destination pair are set to zero since travel on these routes is not possible for the drivers. Thus, the $\mathbf{S}$ matrix can be rather sparse.

Given the driver's route entering the system at a specific time, we need to model the location of the individual in the upcoming times. To model the drivers' location in the system, we use the model developed by [51] in which the drivers' location is computed in a probabilistic fashion. This model can be presented by a matrix $\mathbf{R} \in [0,1]^{(|\mathcal{E}|\cdot|\mathbf{T}|)\times|\mathcal{R}|}$ which estimates the probability of a driver being on a certain link at a given future time, under the assumption that they choose a specific route. Multiple ways to estimate matrix $\mathbf{R}$ are suggested in [51], including an approach based on the use of historical data. In our experiments in subsection V-A, matrix $\mathbf{R}$ is computed based on the volume at the UE state of the system. Given matrix $\mathbf{R}$, it is easy to see that the vector

$$\hat{\mathbf{v}} = \mathbf{RS1} \in \mathbb{R}^{|\mathcal{E}|\cdot|\mathbf{T}|} \tag{3}$$

contains the expected number of vehicles in all the links at each time. Plugging the expression of $\hat{\mathbf{v}}$ in (1), we get the total travel time of the system as

$$\begin{aligned} F(\hat{\mathbf{v}}) &= \sum_{\ell=1}^{|\mathcal{E}|}\sum_{t=1}^{|\mathbf{T}|}(\mathbf{RS1})_{\ell,t}\theta((\mathbf{RS1})_{\ell,t}) \\ &= \sum_{\ell=1}^{|\mathcal{E}|}\sum_{t=1}^{|\mathbf{T}|}(\mathbf{r}_{\ell,t}\mathbf{S1})\theta(\mathbf{r}_{\ell,t}\mathbf{S1}) \end{aligned} \tag{4}$$

where $\mathbf{r}_{\ell,t}$ is the row of matrix $\mathbf{R}$ which corresponds to link ℓ at time t.

To reduce the total travel time of the system, some drivers can be deviated to alternative routes to lower the traffic flow of the congested links. To change the routing assignment of drivers, we need to offer incentives to their organizations such that it can compensate the organizations' financial loss caused by accepting our assignment. For simplicity, we use the total travel time increase of the organization as a measure of financial loss. Although we have estimated the travel time of the system from equation (4), we need to compute the "route travel times" to be able to compare the amount of change in travel time of each driver after offering incentives. Given the route travel times, we compute the incentives using a model that depends on VOT and the amount of increase in the travel time for each organization. In particular, we assume that, given the route assignment to organization i, the incentive value is

$$c_i = \alpha_i \max\left\{0, \sum_{j\in\mathcal{N}_i}\boldsymbol{\delta}^\top\mathbf{s}_i^j - \gamma_i\right\}, \tag{5}$$

where c_i is the incentive offered to organization i, $\alpha_i \in \mathbb{R}_+$ is VOT for organization i, $\boldsymbol{\delta} \in \mathbb{R}_+^{|\mathcal{R}|\cdot|\mathbf{T}|}$ is the travel time of the route for each driver, and γ_i is the sum of the minimum travel time route of each driver of organization i in the absence of incentivization. The variable α_i is designed based on the VOT specific to organization i. This approach allows the model to adjust the VOT for each organization, accommodating the diverse nature of their operations. $\boldsymbol{\delta}$ and γ_i are computed based on the absence of incentivization. When $\sum_{j\in\mathcal{N}_i}\boldsymbol{\delta}^\top\mathbf{s}_i^j - \gamma_i > 0$, the organization's total travel time has increased compared to the baseline of having no incentive, and hence the system will compensate the organization's loss. On the other hand, when $\sum_{j\in\mathcal{N}_i}\boldsymbol{\delta}^\top\mathbf{s}_i^j - \gamma_i < 0$, the organization's travel time is improved after incentivization, and hence no incentivization is required for this particular organization to participate. The details of our method for computing route travel time vector $\boldsymbol{\delta}$ are described next.

Route travel time vector $\boldsymbol{\delta}$: Estimation of the vector $\boldsymbol{\delta}$ requires the volume on each link, which is derived based on the route assignment of the drivers. Let $\mathbf{S}$ denote the routing decision of the drivers. Given $\mathbf{S}$, we can estimate the volume vector $\mathbf{v}$ using (3). By utilizing the BPR function (2) and the estimated volume vector $\hat{\mathbf{v}}$, we can compute the speed of the links. Given the speed of each link, we can determine the vector $\boldsymbol{\delta}$ that contains the travel time of the routes for different time slots and the vector $\boldsymbol{\eta} \in \mathbb{R}_+^{K\cdot|\mathbf{T}|}$ that contains the travel time of the fastest route for different OD pairs for different times (K represents the total number of OD pairs). To do so, we rely on the method provided by [51] and the routing decision of drivers $\mathbf{S}$ at the UE state of the system. Given the minimum travel time between OD pairs in vector $\boldsymbol{\eta}$, we can compute the minimum travel time of organization i as $\gamma_i = (\mathbf{B}_i\boldsymbol{\eta})^\top\mathbf{1}$ where $\mathbf{B}_i \in \{0,1\}^{|\mathcal{N}_i|\times(K\cdot|\mathbf{T}|)}$ is the matrix of shortest travel time assignment of drivers of organization i. $\mathbf{B}_i\boldsymbol{\eta}$ is the vector of the shortest travel time between the OD pair for each driver, and by summing the elements of this vector, we get γ_i.

Proposed formulation: For minimizing the total travel time of the system via providing incentives to organizations, we need to solve the following optimization problem:

$$\begin{aligned} \min_{\{\mathbf{S}_i,c_i\}_{i=1}^n} \quad & \sum_{\ell=1}^{|\mathcal{E}|}\sum_{t=1}^{|\mathbf{T}|}\hat{v}_{\ell,t}\theta_{\ell,t}(\hat{v}_{\ell,t}) \\ \text{s.t.} \quad & \hat{\mathbf{v}} = \sum_{i=1}^{n}\mathbf{RS}_i\mathbf{1} \\ & \mathbf{DS}_i\mathbf{1} = \mathbf{q}_i, \quad \forall i = 1,2,\ldots,n \\ & \mathbf{S}_i^\top\mathbf{1} = \mathbf{1}, \quad \forall i = 1,2,\ldots,n \\ & \mathbf{S}_i \in \{0,1\}^{(|\mathcal{R}|\cdot|\mathbf{T}|)\times|\mathcal{N}_i|}, \quad \forall i = 1,2,\ldots,n \\ & \mathbf{S}_i^\top\boldsymbol{\delta} \le \mathbf{b}_i \odot \mathbf{B}_i\boldsymbol{\eta}, \quad \forall i = 1,2,\ldots,n \\ & c_i \ge \alpha_i(\boldsymbol{\delta}^\top\mathbf{S}_i\mathbf{1} - \gamma_i), \quad \forall i = 1,2,\ldots,n \\ & c_i \ge 0, \quad \forall i = 1,2,\ldots,n \\ & c_1 + c_2 + \cdots + c_n \le \Omega \end{aligned} \tag{6}$$

where $\hat{v}_{\ell,t}$ is an element of vector $\hat{\mathbf{v}}$ that corresponds to the volume of link ℓ at time t, $c_i \in \mathbb{R}_+$ is the cost of incentive assigned to organization i, $\mathbf{D} \in \{0,1\}^{(K\cdot|\mathbf{T}|)\times(|\mathcal{R}|\cdot|\mathbf{T}|)}$ is the matrix of route assignment of the OD pairs, $\mathbf{b}_i \in \mathbb{R}_+^{|\mathcal{N}_i|}$ denotes the factor by which the travel time of an assigned route can be larger than shortest travel time of the OD pair, $\mathbf{B}_i \in \{0,1\}^{|\mathcal{N}_i|\times(K\cdot|\mathbf{T}|)}$ is the matrix of shortest travel time assignment of drivers of organization i, and $\mathbf{q}_i \in \mathbb{R}^{K\cdot|\mathbf{T}|}$ is the vector of the number of drivers of organization i for each OD pair at different times. If there are drivers in the system that do not work for any organization, we can consider them as a single organization whose decision matrix is initialized and has fixed values such that they are assigned to the fastest route (assuming they always select the shortest route). The same idea can be employed for organizations that are not joining the incentivization platform. The following section provides a detailed explanation of the constraints:

Constraint 1 ($\hat{\mathbf{v}} = \sum_{i=1}^{n} \mathbf{R}\mathbf{S}_i\mathbf{1}$): This constraint is the computation of the volume on each link at different times based on the routing assignments for the organizations.

Constraint 2 ($\mathbf{D}\mathbf{S}_i\mathbf{1} = \mathbf{q}_i$): This constraint ensures that the correct number of drivers are assigned to the routes between OD pairs. $\mathbf{S}_i\mathbf{1}$ represents the number of drivers assigned to the different routes. The matrix $\mathbf{D}$ is utilized to aggregate the count of drivers assigned to various routes within the same OD pair. The vector $\mathbf{q}_i$ represents the actual number of drivers from organization i traveling between these OD pairs, and the product $\mathbf{D}\mathbf{S}_i\mathbf{1}$ is required to equal $\mathbf{q}_i$.

Constraint 3 ($\mathbf{S}_i^\top\mathbf{1} = \mathbf{1}$): This constraint simply states that we can only assign one route to each driver of organization i.

Constraint 4 ($\mathbf{S}_i \in \{0,1\}^{(|\mathcal{R}|\cdot|\mathbf{T}|)\times|\mathcal{N}_i|}$): This constraint enforces a binary framework on our decision variables, where 0 indicates not assigning a route and 1 signifies route assignment.

Constraint 5 ($\mathbf{S}_i^\top\boldsymbol{\delta} \leq \mathbf{b}_i \odot \mathbf{B}_i\boldsymbol{\eta}$): This is our fairness and time delivery constraint. Due to different reasons, such as urgent deliveries by some of the organizations' drivers, they may not accept alternative routes that deviate significantly from the fastest route. Moreover, the platform should consider fairness between different drivers in terms of the amount of deviation from the shortest travel time. The fairness and time delivery constraint bounds the deviation of travel time of the assigned routes from the minimum travel time. $\mathbf{S}_i^\top\boldsymbol{\delta}$ represents the travel time of the assigned routes to drivers of organization i. $\mathbf{b}_i \in \mathbb{R}_+^{|\mathcal{N}_i|}$ denotes the factor by which deviation is allowed for each driver of organization i.

Constraints 6 and 7 ($c_i \geq \alpha_i(\boldsymbol{\delta}^\top\mathbf{S}_i\mathbf{1} - \gamma_i)$ and $c_i \geq 0$): These two constraints guarantee (5).

Constraint 8 ($c_1 + c_2 + \cdots + c_n \leq \Omega$): This represents our budget constraint. The scalar c_i denotes the incentive amount allocated to organization i. Ω signifies the total budget available.

For further elaboration on model 6 and its constraints, an illustrative example is presented in Appendix B in the complete version of the work [49].

IV. Incentivization Algorithm and a Distributed Implementation

Optimization problem (6) is of large size and includes binary variables ($\mathbf{S}_i, \forall i = 1,\dots,n$). Thus, solving it efficiently is a challenging task. In this subsection, we propose an efficient algorithm for solving it. First, we relax the binary constraint $\mathbf{S}_i \in \{0,1\}^{(|\mathcal{R}|\cdot|\mathbf{T}|)\times|\mathcal{N}_i|}$ to convex constraint $\mathbf{S}_i \in [0,1]^{(|\mathcal{R}|\cdot|\mathbf{T}|)\times|\mathcal{N}_i|}$ and we refer to this as the relaxed version of problem (6). The objective function is a summation of monomial functions with positive coefficients. Furthermore, $\theta_{\ell,t}$ is an affine mapping of the optimization variable $\mathbf{S}_i$. Since our domain is the nonnegative orthant and monomials are convex in this domain, the objective function is convex. As the constraints of this problem are convex, the relaxed version of problem (6) becomes a convex optimization problem. Thus, standard solvers such as CVX [52] can be used to solve this problem. However, these solvers have large computational complexity because of utilizing methods such as interior point methods [53] with $O(n^3)$ iteration complexity where n is the number of variables. This computational complexity is not practical for our problem. In what follows, we rely on first-order methods with linear computational complexity in n, which is affordable in our problem. The reformulation is provided in Appendix B. This reformulation is amenable to the Alternating Direction Method of Multiplier (ADMM) method [54–58], which is a first-order method and scalable. Appendix C provides an overview of ADMM, the fundamental component of our framework. The steps of the resulting algorithm are provided in Algorithm 1 in Appendix D. The details of the derivation of this algorithm are provided in Appendix E in the complete version of the work [49]. Due to the distributed setting of Algorithm 1 using the ADMM method, it also provides the potential benefits associated with federated learning and distributed systems [59, 60].

In the relaxed version of problem (6), different solutions $\mathbf{S}_i^*$ with a fixed $\mathbf{S}_i^*\mathbf{1} = \mathbf{u}_i^*$ yield the same objective value if $\mathbf{S}_i^*$ satisfies all the constraints. Thus, potentially infinitely many solutions to our convex problem exist, and many are not binary. To promote a binary solution for the final decision, we introduce the following regularizer into the objective function of the relaxed version of problem (6):

$$\Re(\mathbf{S}) = -\frac{\tilde{\lambda}}{2}\sum_{i=1}^{n}\sum_{j=1}^{|\mathcal{N}_i|}\sum_{r=1}^{|\mathcal{R}|}\sum_{t=1}^{|\mathbf{T}|}(\mathbf{S}_i)_{j,r,t}((\mathbf{S}_i)_{j,r,t} - 1) \tag{7}$$

where $\tilde{\lambda} \in \mathbb{R}_+$ is the regularization parameter and $(\mathbf{S}_i)_{(j,r,t)} \in [0,1]$. This regularizer has the effect of driving the elements of matrix $\mathbf{S}$ towards the binary domain $\{0,1\}$. The regularizer penalizes any deviations from this domain in the objective function. While convexity is sacrificed due to regularization, ADMM can still be convergent in nonconvex problems [57].

Algorithm 1 solves the relaxed version of problem (6). Since the solution to the relaxed version of problem (6) may not be binary (due to relaxation), we need to project it back to the feasible region. For computational purposes, we suggest using ℓ_1 projection by solving the following mixed integer (linear) problem (MILP)

$$\begin{aligned}
\min_{\{\mathbf{S}_i,c_i\}_{i=1}^n} \quad & \sum_{i=1}^{n}\|\mathbf{S}_i\mathbf{1} - \mathbf{u}_i^*\|_1 \\
\text{s.t.} \quad & \mathbf{D}\mathbf{S}_i\mathbf{1} = \mathbf{q}_i, \quad \forall i = 1,2,\dots,n \\
& \mathbf{S}_i^\top\mathbf{1} = \mathbf{1}, \quad \forall i = 1,2,\dots,n \\
& \mathbf{S}_i \in \{0,1\}^{(|\mathcal{R}|\cdot|\mathbf{T}|)\times|\mathcal{N}_i|}, \quad \forall i = 1,2,\dots,n \\
& \mathbf{S}_i^\top\boldsymbol{\delta} \leq \mathbf{b}_i \odot \mathbf{B}_i\boldsymbol{\eta}, \quad \forall i = 1,2,\dots,n \\
& c_i \geq \alpha_i(\boldsymbol{\delta}^\top\mathbf{S}_i\mathbf{1} - \gamma_i), \quad \forall i = 1,2,\dots,n \\
& c_i \geq 0, \quad \forall i = 1,2,\dots,n \\
& c_1 + c_2 + \cdots + c_n \leq \Omega
\end{aligned} \tag{8}$$

where $\mathbf{u}_i^*, \forall i = 1,2,\dots,n$ is the optimal solution obtained by Algorithm 1. Clearly, this problem can be reformulated as a MILP problem and solved using off-the-shelf solvers like

Gurobi. Solving problem (8) can be easier than problem (6). Problems (6) and (8) have the same variable size and similar constraints, but the objective functions are different. While the objective function in problem (8) can be restructured as a linear programming problem, we have a polynomial objective function in problem (6) that introduces more complexity.

V. Experiments

We evaluate our incentive scheme's effectiveness using Los Angeles area data. The presence of multiple routes between most OD pairs makes the Los Angeles area particularly suitable for our assessment. We use the data collected by the Archived Data Management System (ADMS), a comprehensive transportation dataset compilation by University of Southern California researchers [61]. This system aggregates data from Los Angeles, Orange, San Bernardino, Riverside, and Ventura Counties, offering a robust data source for analysis.

For our evaluations, we need to estimate the OD matrix. The (i, j)-th entry of the OD matrix represents the count of drivers traveling between origin i and destination j. We need to estimate the OD matrix using the available network flow information due to drivers' routing data unavailability. The OD matrix estimation problem is challenging due to its under-determined nature [62–64]. OD matrices are categorized as static or dynamic [65]. However, many dynamic OD estimation (DODE) methods are computationally impractical for our high-resolution data. Additionally, some studies rely on existing OD matrix data [66–69], which we lack. Given these constraints, we adopt the OD estimation algorithm proposed by [51]. Note that OD estimations in our study serve as an input to our incentivization model rather than being the focus of our analysis, as we do not propose a new OD estimation algorithm.

The base VOT of our experiments is derived from the estimation by [70], which is \$2.63 per minute or \$157.8 per hour. In our experiments, we apply a uniform VOT across all organizations. We note that, in practice, we do not initially know the exact VOT of passengers and drivers. Moreover, the perceived VOT by organizations can change over time because the incentive policy would necessitate algorithmic adjustments within the organizations. Specifically, they would need to modify their payment algorithms to allocate the received incentives between passengers and drivers that accept longer routes. Such algorithmic updates would allow the organizations to optimize their operations and services in response to the incentive policy. The incentivization platform can learn the VOT of passengers and drivers by continuously observing their acceptance/rejection behavior through techniques in online/reinforcement learning. Learning the VOT is beyond the scope of our work, and we assume the VOT is known. All the codes are publicly available at: https://github.com/ghafeleb/Incentive_Systems_for_New-_Mobility_Services.

A. Simulation Model

First, we extract sensor details, including their locations. We extract the speed and volume data of selected sensors. Nodes for the network graph are chosen from on-ramps and highway intersections. Connecting link data is derived from in-between sensors. Node distances are determined via Google Maps API. Data preparation workflow is shown in Fig. 4. The network under consideration includes highways surrounding Downtown Los Angeles, as depicted in Fig. 5, and consists of 12 nodes, 32 links, and a total road length of 288.1 miles. We have 144 OD pairs, and we employ the algorithm from [51] on the network's speed and volume data for OD estimation. Fig. 6 shows the total estimated incoming drivers per time interval. We explore 3 routing options for each OD pair. Initially, the shortest path is determined. Subsequently, links in the first path are removed to uncover the second shortest path if available. This process is repeated for the third route. Based on this process, we find 270 paths between OD pairs.

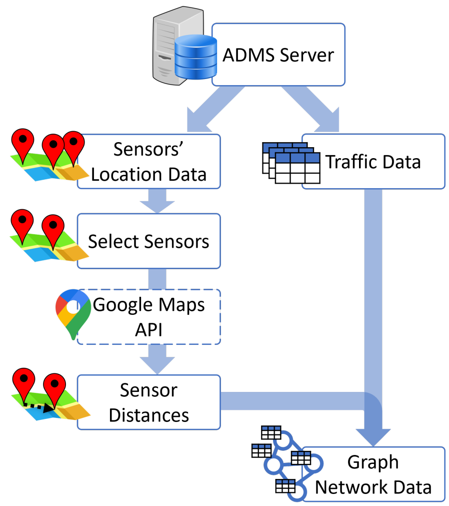

Fig. 4. Data preparation workflow. First, traffic data and sensors' location data are received from the ADMS Server. Next, sensors' location data is processed to compute sensor distances. Finally, sensor distances and traffic data are combined to create the graph network data.

In practice, the prediction of travel time and OD estimations are handled by organizations using their sophisticated software and data-collecting tools. By incorporating these prediction tools, the framework can consider external factors such as weather conditions in traffic predictions or road closures in finding possible routes because our approach focuses on short-term planning (only a few hours ahead).

We focus on incentivizing the organizations to change their behavior for the 7 AM to 8 AM interval (which is the rush hour based on the estimated number of incoming drivers in Fig. 6). Although we have selected 7 AM to 8 AM as the incentivization time period, we also include 8 AM to 8:30 AM in our experiments because some of the drivers entering between 7 AM and 8 AM may not finish their route before 8 AM. To track the effect of these drivers on the total travel time of the system, we include traffic flow from 8 AM to 8:30 AM in our analysis as well. The OD estimation algorithm's projected total count of drivers entering the system from 6 AM to 9 AM is illustrated in Fig. 6. From 7 AM to 8:30 AM, a total of 11985 drivers enter the system.

We consider the traffic volume of the network at UE in our

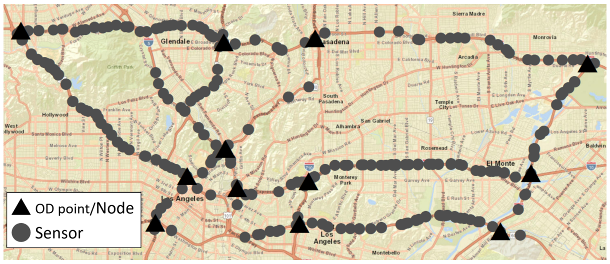

Fig. 5. Studied region and the highway sensors inside the region. This region encompasses several areas notorious for high traffic congestion, particularly Downtown Los Angeles.

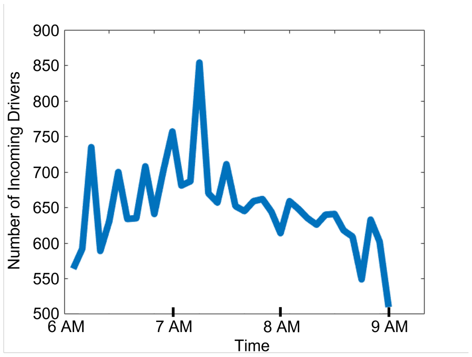

Fig. 6. Total estimated number of drivers entering the system over time (in 5-minute intervals). The plot shows that traffic peak happens between 7 AM and 8 AM.

baseline. To compute the volume of the network at UE, we use the UE algorithm in [48]. The algorithm receives the volume (historical data) and OD estimation as inputs and returns the matrix $\mathbf{R}_{\text{UE}}$ and route travel time vector $\boldsymbol{\delta}_{\text{UE}}$ at UE. To compute the cost of organizations' incentivization, we need to know the route travel times when drivers have made decisions based on the UE route travel time $\boldsymbol{\delta}_{\text{UE}}$. Hence, we compute the new volume vector $\mathbf{v}_{\text{new}} = \mathbf{R}_{\text{UE}}\mathbf{S}_{\text{UE}}\mathbf{1}$ where $\mathbf{S}_{\text{UE}}$ is the assignment of drivers to the fastest route based on the UE route travel time vector $\boldsymbol{\delta}_{\text{UE}}$. Using the BPR function, volume vector $\mathbf{v}_{\text{new}}$, and $\boldsymbol{\delta}_{\text{UE}}$, we compute $\boldsymbol{\delta}$ that denotes the travel time of the routes if drivers make decision based on $\boldsymbol{\delta}_{\text{UE}}$ and $\boldsymbol{\eta}$ denotes the minimum travel time between the different OD pairs.

B. Results

In this subsection, using our model and algorithm, we study the impact of organization incentivization for different budget values, the number of organizations, VOTs, and the percentage of drivers who are employed by the organizations in the incentivization program. The remaining drivers are assumed to be background drivers who follow the $\boldsymbol{\delta}_{\text{UE}}$. We consider four scenarios for the percentage of drivers that enter the system between 7 AM and 8 AM and belong to organizations that we can incentivize (penetration rate): I. 5% (407 drivers) II. 10% (812 drivers) III. 15% (1221 drivers) IV. 20% (1626 drivers). Selected drivers in each scenario are included in scenarios with larger user percentages to have a standard comparison between scenarios. Drivers in each organization are selected uniformly at random.

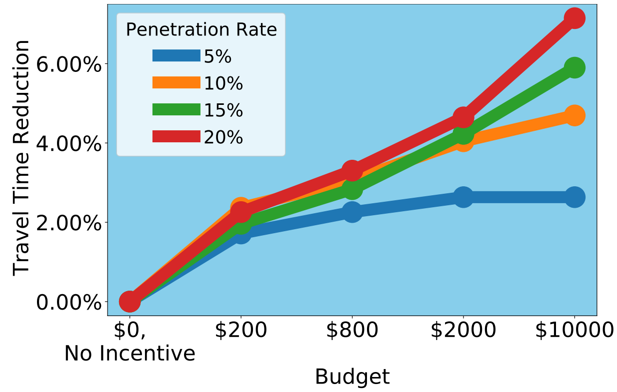

Fig. 7. Percentage of travel time decrease with different budgets at VOT=$157.8 per hour using Algorithm 1. The amount of travel time reduction shows a positive correlation with the amount of incentivization budget and the penetration rate

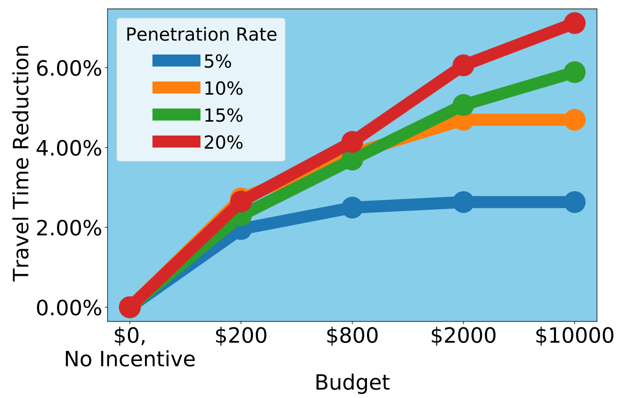

Fig. 8. Percentage of travel time decrease with different budgets at VOT=$78.9 per hour using Algorithm 1. The amount of travel time decrease is similar or larger compared to Fig. 7 with VOT=$157.8 due to the smaller cost of incentivization.

The percentage of travel time decrease with incentivization as compared to a system with no incentivization scheme with VOT of \$157.8 is presented in Fig. 7 for different penetration rates. In our plots, the budget of \$0 shows the case of a no-incentivization platform. We observe that by increasing the budget, the amount of decrease in travel time increases (as expected). This decrease is more for the same budgets at larger penetration rates because the model has access to more drivers to select and has more flexibility to recommend alternative routes. For the purpose of sensitivity analysis, we also provide travel time decrease for all penetration rates with a different VOT of $\frac{\$157.8}{2} = \$78.9$ per hour in Fig. 8. The comparison of results for different VOTs in Fig. 7 and Fig. 8 shows that for a very large budget, the decrease in travel time is almost similar. This is because none of the models utilize the entire budget at a \$10000 budget. However, when budgets are limited, the performance disparity can increase up to 1.41% due to lower incentivization costs associated with the smaller VOT.

For the next analyses of our numerical results, we only report the results for our base VOT (\$2.63 per minute or \$157.8 per hour) because the results follow similar patterns with VOT of \$78.9. In Fig. 9, we present the total incentivization cost for different budgets and penetration rates when there is

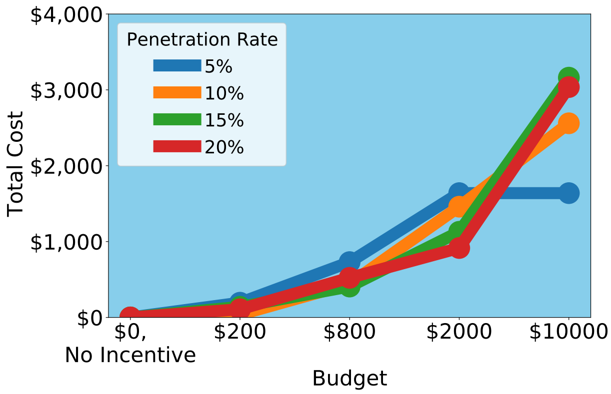

Fig. 9. Total cost of incentivization of one organization with different budgets and different penetration rates at VOT=$157.8 per hour using Algorithm 1. Incentivization cost increases with the amount of budget as the model incentivizes more drivers to reduce traffic. The cost is larger at larger penetration rates at $10000 budget because the model incentivizes more drivers. At smaller budgets, the incentivization cost is smaller at larger penetration rates because of more flexibility in selecting drivers at limited budgets.

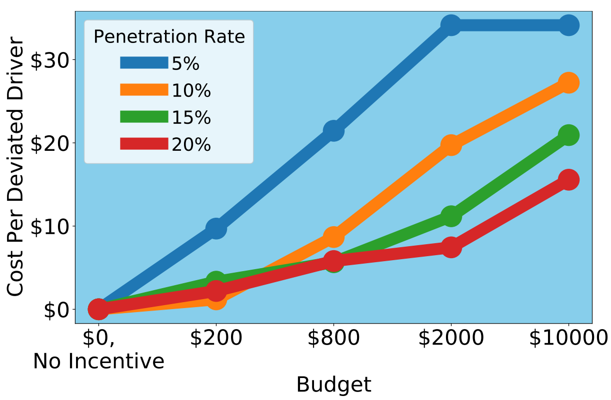

Fig. 10. Cost of incentivization per deviated drivers of one organization with different budgets and different penetration rates at VOT=$157.8 per hour using Algorithm 1. At larger penetration rates, the platform can incentivize drivers more efficiently due to access to a larger pool. The platform spends smaller incentivization amount per deviated driver at larger penetration rates.

one organization in the system. This cost increases when the available budget is more. This pattern shows that the platform can utilize the resources when it has access to more money. We observe that more involvement of drivers at $800 and $2000 budget leads to a slightly smaller cost at larger penetration rates because of more flexibility in selecting drivers. At a $10000 budget, the platform does not exhaust the whole budget at any penetration rate. Hence, it spends more on incentivization at larger penetration rates by incentivizing a greater number of participants. Fig. 10 shows the cost per deviated driver. The cost per driver is significantly smaller in larger penetration rates because the model has more flexibility in choosing the drivers efficiently. Moreover, the cost per driver increases with the budget. This shows that our model utilizes our budget efficiently by providing more affordable incentives first when the budget is low. As TABLE I shows, the number of incentivized drivers in larger penetration rates is larger because there are more drivers for selection.

The number of organizations in the system can alter the total travel time and cost. Fig. 11 illustrates the percentage decrease of travel time and total cost when there are different number of organizations in the system at a 5% penetration rate. As an extreme case, we also include the case that each organization contains one driver (i.e., we incentivize individuals rather than organizations). In Fig. 11, we observe a larger cost for reducing the same amount of travel time decrease when there are more organizations in the system. The intuitive reason behind this observation is as follows. For each organization, after incentivization, some drivers lose time, and some gain travel time. At the organizational level, the time changes of drivers can cancel each other out, and hence we may not need to compensate the organization significantly. When the number of drivers per organization decreases, the canceling effect becomes weaker, and the incentivization costs more. This is in line with our discussion in Section II. This also explains why incentivizing organizations is much more cost-efficient than incentivizing individual drivers. This observation remains consistent across other penetration rates; therefore, corresponding plots for other rates are not provided.

Our experiments use Algorithm 1 to solve the relaxed version of problem (6) and utilizes the Gurobi solver to solve the MILP problem (8). We compare our approach against solving the MIP problem (6), utilizing Gurobi and MOSEK. These solvers are recognized as state-of-the-art, off-the-shelf commercial tools for linear and mixed integer optimization problems. We configure the relative mixed integer programming optimality gap at 0.01 for both solvers to ensure an optimal trade-off between computational speed and accuracy. Fig. 12

Penetration Rate	Budget			
	$200	$800	$2000	$10000
5%	20	34	48	48
10%	31	51	74	94
15%	42	72	101	151
20%	49	90	123	195

TABLE I
DISTRIBUTION OF THE NUMBER OF DRIVERS THAT WERE ASSIGNED TO AN ALTERNATIVE ROUTE USING ALGORITHM 1.

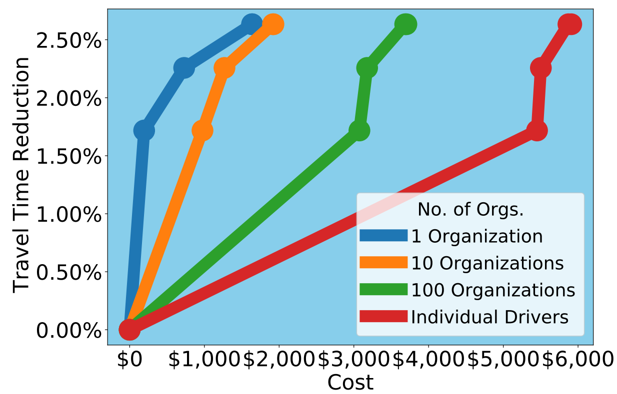

Fig. 11. Travel time decrease vs. incentivization cost for different number of organizations at a 5% penetration rate and VOT=$157.8 per hour using Algorithm 1. The incentivization cost for the same travel time reduction is smaller when the number of organizations is smaller. This phenomenon is due to the cancel-out effect between the gain and loss of drivers of the organizations.

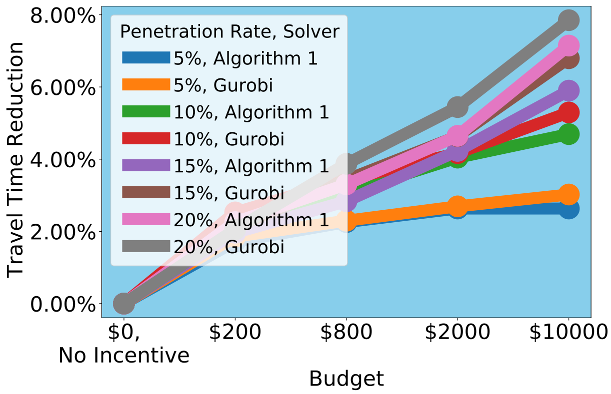

Fig. 12. Comparison of travel time reduction percentage using different solvers with different penetration rates and budgets at VOT=$157.8 per hour. Gurobi exhibits a slight performance advantage over Algorithm 1 at higher penetration rates and budgets.

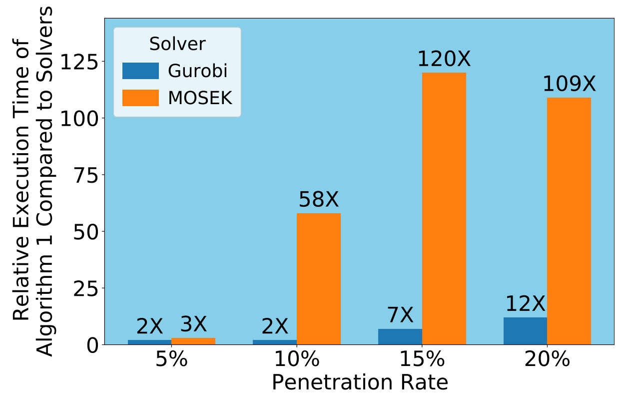

Fig. 13. Comparison of the relative execution time of Algorithm 1 vs. different solvers at different penetration rates at VOT=$157.8 per hour. Algorithm 1 execution time consistently outperforms Gurobi and MOSEK up to 12 and 120 times, respectively.

shows that the Gurobi solver has a slightly better travel time reduction compared to our method. MOSEK is not included in this plot because its performance closely mirrors that of Gurobi. Although the solvers show a slight advantage in reducing travel time, our presented method significantly outpaces these solvers when parallel computation techniques are applied. As Fig. 13 shows, our method achieves speeds up to 12 times faster than Gurobi and 120 times faster than MOSEK, demonstrating a considerable advantage in computational efficiency. This enhanced speed does not only translate to quicker solutions but also suggests potential for real-time application in dynamic traffic management scenarios where rapid decision-making is critical. Moreover, Fig. 14 illustrates that, at $10000 budget, the Gurobi solver spends significantly more (up to $5000) on incentivization compared to Algorithm 1. This discrepancy highlights the cost-efficiency of our algorithm, particularly in managing budget allocations effectively while achieving comparable traffic management outcomes. The potential reason is that Gurobi employs branch-and-bound and linear programming solvers to find the solution in a finite number of steps, relying on extreme points. In contrast, Algorithm 1 is based on a first-order method and asymptotically converges to the solution, stopping upon finding an ϵ-optimal solution. Due to the similar incentivization cost of MOSEK and Gurobi, a comparative analysis for MOSEK is not included.

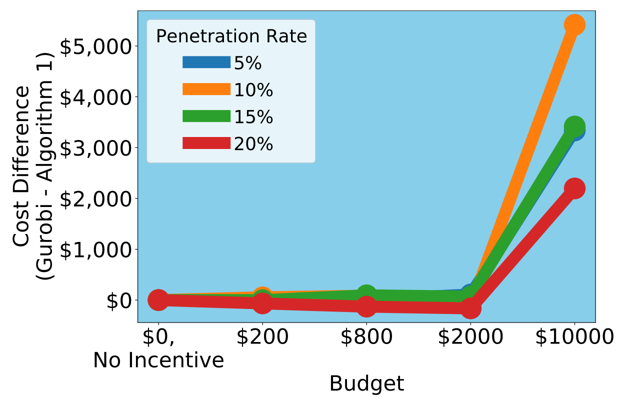

Fig. 14. Comparison of the relative incentivization cost using Algorithm 1 vs. Gurobi at different penetration rates, VOT=$157.8 per hour, and one organization. Both methods utilize similar incentivization amount for smaller budgets but at $10000 budget, Gurobi spends up to $5000 more.

VI. Conclusion

In this paper, we study the problem of incentivizing organizations to reduce traffic congestion. To this end, we developed a mathematical model and provided an algorithm for offering organization-level incentives. In our framework, a central planner collects the origin-destination and routing information of the organizations. Then, the central planner utilizes this information to offer incentive packages to organizations to incentivize a system-level optimal routing strategy. In particular, we focused on minimizing the total travel time of the network. However, other utilities can be used in our framework. Finally, we employed data from the Archived Data Management System (ADMS) to evaluate the performance of our model and algorithm in a representative traffic scenario in the Los Angeles area. A 6.90% reduction in the total travel time of the network was reached by our framework in the experiments. More importantly, we observed that incentivizing companies/organizations is more cost-efficient than incentivizing individual drivers. As future work, it is important to study the effect of incentivization to change the start time of the trip. This is particularly relevant in future mobility services because many of them, such as delivery services, are flexible in terms of trip time to a certain degree. In addition, we can consider the stochastic nature of making decisions in routing by individual drivers. Moreover, we can extend the incentivization framework to the case that not all organizations accept their received offer. Our platform also has the limitation of assuming VOT is given and fixed. Furthermore, as a potential legal, ethical, and practical constraints, we should consider the privacy of individuals' data. We can adopt approaches similar to those used in previous incentivization projects with real-world implementations [37–40] to address this issue. Further discussions on limitations and scalability of the platform are provided in Appendix E.

APPENDIX A
LIST OF NOTATIONS

Traffic network spatiotemporal parameters:

- $\mathcal{G}$: Directed graph of the traffic network
- $\mathcal{V}$: Set of nodes of graph $\mathcal{G}$ which correspond to major intersections and ramps
- $\mathcal{E}$: Set of edges of graph $\mathcal{G}$ which correspond to the set of road segments
- $|\mathcal{E}|$: Total number of road segments/edges in the network $\mathcal{G}$ (i.e. the cardinality of the set $\mathcal{E}$)
- ℓ: An edge of graph $\mathcal{G}$ which corresponds to a link/road segment in the traffic network
- $\mathcal{R}_j$: Set of possible route options for driver j
- $\mathcal{R}$: Total set of possible route options for all OD pairs
- $|\mathcal{R}|$: Total number of possible route options (i.e. the cardinality of the set $\mathcal{R}$)
- $\mathbf{r}$: Route vector
- $\mathbf{T}$: Set of time of periods
- $|\mathbf{T}|$: Number of time units (i.e. the cardinality of $\mathbf{T}$)
- $\theta_{\ell,t}$: Travel time of link ℓ at time t
- $F(.)$: Total travel time function
- $T_{\mathbf{r}}$: The travel time for route $\mathbf{r}$

BPR function and its parameters:

- $f_{\text{BPR}}(.)$: BPR function
- v: The traffic volume of the link
- w: The practical capacity of the link
- θ_0: The free flow travel time of the link

Optimization model parameters:

- $\mathcal{N}_i$: Set of drivers of organization i
- $|\mathcal{N}_i|$: Total number of drivers of organization i (i.e. the cardinality of set $\mathcal{N}_i$)
- $\mathcal{N}$: Set of all drivers
- $|\mathcal{N}|$: Total number of drivers (i.e., the cardinality of set $\mathcal{N}$)
- $\mathbf{v}_t$: Volume vector of road segments at time t
- $\hat{\mathbf{v}}$: The vector of the estimated volume of links at different times in the horizon
- $\hat{v}_{\ell,t}$: The $(|\mathcal{E}| \times t + \ell)^{th}$ element of vector $\hat{\mathbf{v}}$ representing the volume of the ℓ^{th} link at time t
- $\mathbf{R}$: The matrix of the probability of a driver being at each link given their route
- $\mathbf{r}_{\ell,t}$: The row of matrix $\mathbf{R}$ that corresponds to link ℓ at time t
- $\mathbf{D}$: The matrix of route assignments of the OD pairs
- $\mathbf{q}_i$: The vector of number of drivers of organization i for each OD pair
- $\boldsymbol{\delta}$: The vector of travel time of routes at different times
- $\boldsymbol{\eta}$: The vector of shortest travel time between different OD pairs at different times
- $\mathbf{b}_i$: This vector contains the factors by which the travel time of assigned routes can be larger than the shortest travel time of the drivers of organization i
- $\mathbf{B}_i$: The matrix of shortest travel time assignment of drivers of organization i
- α_i: VOT for organization i
- $\boldsymbol{\alpha}$: The vector of VOT values for the different organizations
- γ_i: Total travel time of organization i in the absence of incentivization platform
- Ω: Budget for incentivization
- K: The number of OD pairs

Decision variables:

- $s_i^{\mathbf{r},j}$: Decision parameter indicates whether route $\mathbf{r}$ is assigned to driver j from organization i
- $\mathbf{s}_i^j$: The binary route assignment vector of driver j from organization i
- $\mathbf{S}_i$: Decision matrix of drivers of organization i
- $\mathbf{S}$: Decision matrix of all drivers
- c_i: The cost of incentive offered to organization i

APPENDIX B
REFORMULATED OPTIMIZATION MODEL FOR THE ADMM ALGORITHM

To solve the relaxed version of problem (6) efficiently, we present a distributed algorithm based on this reformulation

$$
\begin{aligned}
\min_{\substack{\mathbf{S},\mathbf{H},\mathbf{W},\mathbf{Z},\mathbf{u},\boldsymbol{\beta},\\ \omega,\mu_i,\tilde{\beta},\mathbf{c}}} \quad & \sum_{\ell=1}^{|\mathcal{E}|}\sum_{t=1}^{|\mathbf{T}|} \hat{v}_{\ell,t}\theta_{\ell,t}(\hat{v}\ell,t) \\
& - \frac{\tilde{\lambda}}{2}\sum_{r=1}^{\mathcal{R}}\sum_{t=1}^{|\mathbf{T}|}\sum_{i=1}^{n}(\mathbf{Z}_i)_{r,t}((\mathbf{Z}_i)_{r,t}-1) \\
\text{s.t.} \quad & \mathbf{S}_i\mathbf{1} = \mathbf{u}_i, \quad \forall i = 1,2,\ldots,n \\
& \boldsymbol{\omega} = \tilde{\mathbf{R}}\mathbf{u} \\
& \tilde{\mathbf{D}}\mathbf{u} = \mathbf{q}, \quad \forall i = 1,2,\ldots,n \\
& \mathbf{W}_i^\top\mathbf{1} = \mathbf{1}, \quad \forall i = 1,2,\ldots,n \\
& \mathbf{S}_i = \mathbf{W}_i, \quad \forall i = 1,2,\ldots,n \\
& \mathbf{H}_i^\top\boldsymbol{\delta} + \boldsymbol{\beta}_i = \mathbf{b}_i \odot \mathbf{B}_i\boldsymbol{\eta}, \quad \forall i = 1,2,\ldots,n \\
& \mathbf{S}_i = \mathbf{H}_i, \quad \forall i = 1,2,\ldots,n \\
& \boldsymbol{\beta}_i \geq 0, \quad \forall i = 1,2,\ldots,n \\
& \mathbf{Z}_i \in [0,1]^{(|\mathcal{R}|\cdot|\mathbf{T}|)\times(|\mathcal{N}_i|)}, \quad \forall i = 1,2,\ldots,n \\
& \tilde{\mathbf{I}}\tilde{\mathbf{c}} = \boldsymbol{\alpha}\odot(\boldsymbol{\Delta}\mathbf{u}-\boldsymbol{\gamma}) \\
& \tilde{\mathbf{c}} \geq 0, \quad \tilde{\mathbf{c}}^\top\tilde{\mathbf{1}} + \tilde{\beta} = \Omega, \quad \tilde{\beta} \geq 0 \\
& \mathbf{S}_i = \mathbf{Z}_i, \quad \forall i = 1,2,\ldots,n,
\end{aligned} \tag{9}
$$

where $\mathbf{S} = \{\mathbf{S}_i\}_{i=1}^n$, $\mathbf{H} = \{\mathbf{H}_i\}_{i=1}^n$, $\mathbf{W} = \{\mathbf{W}_i\}_{i=1}^n$, $\mathbf{Z} = \{\mathbf{Z}_i\}_{i=1}^n$, $\mathbf{u} = \{\mathbf{u}_i\}_{i=1}^n$, and $\boldsymbol{\beta} = \{\boldsymbol{\beta}_i\}_{i=1}^n$.

APPENDIX C
REVIEW OF ALTERNATING DIRECTION METHOD OF MULTIPLIERS (ADMM)

In this section, we review the Alternating Direction Method of Multipliers (ADMM), which is the main building block of our framework. ADMM developed in [55] and [56] aims at solving linearly constrained optimization problems of the form

$$\min_{w,z} \quad h(w) + g(z) \qquad \text{s.t.} \quad Aw + Bz = c,$$

where $w \in \mathbb{R}^{d_1}$, $z \in \mathbb{R}^{d_2}$, $c \in \mathbb{R}^k$, $A \in \mathbb{R}^{k\times d_1}$, and $B \in \mathbb{R}^{k\times d_2}$. By forming the augmented Lagrangian function

$$\mathcal{L}(w,z,\lambda) \triangleq h(w)+g(z)+\langle\lambda, Aw+Bz-c\rangle+\frac{\rho}{2}\|Aw+Bz-c\|_2^2,$$

each iteration of ADMM applies alternating minimization to the primal variables and gradient ascent to the dual variables:

$$\begin{aligned} \text{Primal Update:} \quad & w^{r+1} = \arg\min_{w} \ \mathcal{L}(w, z^r, \lambda^r), \\ & z^{r+1} = \arg\min_{z} \ \mathcal{L}(w^{r+1}, z, \lambda^r) \\ \text{Dual Update:} \quad & \lambda^{r+1} = \lambda^r + \rho\left(Aw^{r+1} + Bz^{r+1} - c\right) \end{aligned} \tag{10}$$

This algorithm is extensively explored in the optimization literature (see [54] for a monograph on the use of this algorithm in convex distributed optimization and [57] for its use in non-convex continuous optimization). If we apply ADMM to (9), we get Algorithm 1. See the details of the derivation of our proposed algorithm in Appendix E in the complete version of the work [49].

Appendix D
Distributed Incentivization Algorithm

Algorithm 1 solves the relaxed version of problem (6). In this algorithm, we use the projection operator $\Pi(\cdot)_{[0,1]}$ that projects elements of a matrix to the interval $[0, 1]$. $\Pi(\cdot)_{\mathbb{R}_+}$ is also a projection operator but projects elements of a matrix to $\mathbb{R}_+$. Notice that in Algorithm 1, the computation load of steps 9, 15, 16, and 17 is extensive because matrices $\mathbf{S}, \mathbf{W}, \mathbf{H}$ and $\mathbf{Z}$ have large sizes. However, each column in these matrices corresponds to one driver and these steps are not coupled so we can perform the computation of each column in parallel by leveraging parallel computation. The notations used in Algorithm 1 are defined below.

$$\boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix} \qquad \mathbf{q} = \begin{bmatrix} \mathbf{q}_1 \\ \vdots \\ \mathbf{q}_n \end{bmatrix} \qquad \boldsymbol{\lambda}_i = \begin{bmatrix} \boldsymbol{\lambda}_{i,1} \\ \vdots \\ \boldsymbol{\lambda}_{i,n} \end{bmatrix}, i = 1, 3$$

$$\tilde{\mathbf{R}} = [\mathbf{R} \quad \dots \quad \mathbf{R}] \qquad \tilde{\mathbf{I}} = [\mathbf{I} \quad -\mathbf{I}]$$

$$\tilde{\boldsymbol{\alpha}} = \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{bmatrix} \qquad \boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_n \end{bmatrix} \qquad \tilde{\mathbf{u}}^t = \begin{bmatrix} \mathbf{S}_1^t \mathbf{1} \\ \vdots \\ \mathbf{S}_n^t \mathbf{1} \end{bmatrix}$$

$$\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} & & \\ & \ddots & \\ & & \mathbf{D} \end{bmatrix} \quad \boldsymbol{\Delta} = \begin{bmatrix} \boldsymbol{\delta} & & \\ & \ddots & \\ & & \boldsymbol{\delta} \end{bmatrix} \quad \tilde{\mathbf{1}} = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} \quad \tilde{\mathbf{c}} = \begin{bmatrix} \mathbf{c} \\ \boldsymbol{\mu} \end{bmatrix}$$

Appendix E
Limitations and Further Discussions

While our platform demonstrates significant potential, several limitations and considerations warrant further discussion: First, our simulations assume that VOT is given and fixed. Although these values can be learned by observing drivers' and passengers' behavior, learning VOT is beyond the scope of our work and we assumed it is known. Moreover, the BPR function used in our simulations to compute travel time can sometimes be inaccurate. However, our modular design allows for any non-linear travel time computation function, offering flexibility in practice. A potential practical limitation

Algorithm 1 Distributed Organization-Level Incentivization via ADMM

1: **Input**: Initial values: $\boldsymbol{\omega}^0$, $\mathbf{S}_i^0$, $\mathbf{H}_i^0$, $\mathbf{W}_i^0$, $\mathbf{Z}_i^0$, $\mathbf{u}^0$, β_i^0, $\tilde{\beta}^0$, $\tilde{\mathbf{c}}^0$, $\boldsymbol{\lambda}_{1,i}^0 \in \mathbb{R}^{|\mathcal{R}|\cdot|\mathbf{T}|}$, $\boldsymbol{\lambda}_2^0 \in \mathbb{R}^{|\mathcal{E}|\cdot|\mathbf{T}|}$, $\boldsymbol{\lambda}_{3,i}^0 \in \mathbb{R}^{K\cdot|\mathbf{T}|}$, $\boldsymbol{\lambda}_{4,i}^0 \in \mathbb{R}^{|\mathcal{R}|\cdot|\mathbf{T}|}$, $\boldsymbol{\Lambda}_{5,i}^0 \in \mathbb{R}^{(|\mathcal{R}|\cdot|\mathbf{T}|)\times|\mathcal{N}_i|}$, $\boldsymbol{\lambda}_{6,i}^0 \in \mathbb{R}^{|\mathcal{N}_i|}$, $\boldsymbol{\lambda}_7^0 \in \mathbb{R}^n$, $\boldsymbol{\Lambda}_{8,i}^0 \in \mathbb{R}^{(|\mathcal{R}|\cdot|\mathbf{T}|)\times|\mathcal{N}_i|}$, $\lambda_9^0 \in \mathbb{R}$, $\boldsymbol{\Lambda}_{10,i}^0 \in \mathbb{R}^{(|\mathcal{R}|\cdot|\mathbf{T}|)\times|\mathcal{N}_i|}$, Dual update step: ρ, Number of iterations: $\tilde{T}$.
2: **for** $t = 0, 1, \dots, \tilde{T}$ **do**
3: **for** $\ell = 0, 1, \dots, |\mathcal{E}|$ **do**
4: **for** $\hat{t} = 1, \dots, |\mathbf{T}|$ **do**
5: $\omega_{\ell,\hat{t}}^{t+1} = \operatorname*{argmin}_{\omega_{\ell,\hat{t}}} \omega_{\ell,\hat{t}}\theta_{\ell,t}(\omega_{\ell,\hat{t}}) + \boldsymbol{\lambda}_{2,(\ell,\hat{t})}^t(\omega_{\ell,\hat{t}} - \mathbf{r}_{\ell,\hat{t}}\,(\sum_{i=1}^n \mathbf{u}_i^t)) + \frac{\rho}{2}(\omega_{\ell,\hat{t}} - \mathbf{R}_{\ell,\hat{t}}\,(\sum_{i=1}^n \mathbf{u}_i^t))^2$
6: **end for**
7: **end for**
8: **for** $i = 1, \dots, n$ **do**
9: $\mathbf{S}_i^{t+1} = (-\boldsymbol{\lambda}_{1,i}^t \mathbf{1}^\top - \boldsymbol{\Lambda}_{5,i}^t - \boldsymbol{\Lambda}_{8,i}^t - \boldsymbol{\Lambda}_{10,i}^t + \rho \mathbf{u}_i^t \mathbf{1}^{t\top} + \rho \mathbf{W}_i^t + \rho \mathbf{H}_i^t + \rho \mathbf{Z}_i^t)(\rho \mathbf{1}\mathbf{1}^\top + 3\rho \mathbf{I})^{-1}$
10: $\beta_i^{t+1} = \Pi\left(\frac{1}{\rho}(-\boldsymbol{\lambda}_{6,i}^t - \rho \mathbf{H}_i^{t\top}\boldsymbol{\delta} + \rho \mathbf{b}_i \odot (\mathbf{B}_i \boldsymbol{\eta}))\right)_{\mathbb{R}_+}$
11: **end for**
12: $\tilde{\mathbf{c}}^{t+1} = \Pi(\frac{1}{\rho}(\tilde{\mathbf{I}}^\top \tilde{\mathbf{I}} + \tilde{\mathbf{1}}\tilde{\mathbf{1}}^\top)^{-1}(\tilde{\mathbf{I}}^\top \boldsymbol{\lambda}_7^t - \lambda_9^t \tilde{\mathbf{1}} - \rho \tilde{\mathbf{I}}^\top (\boldsymbol{\alpha} \odot \boldsymbol{\gamma}) + \rho \tilde{\mathbf{I}}^\top (\boldsymbol{\alpha} \odot (\boldsymbol{\Delta}^\top \mathbf{u}^t)) - \rho \tilde{\boldsymbol{\beta}} \tilde{\mathbf{1}} + \rho \Omega \tilde{\mathbf{1}})_{\mathbb{R}_+}$
13: $\mathbf{u}^{t+1} = \frac{1}{\rho}(\mathbf{I} + \tilde{\mathbf{R}}^\top \tilde{\mathbf{R}} + \tilde{\mathbf{D}}^\top \tilde{\mathbf{D}} + (\Delta\tilde{\alpha})(\Delta\tilde{\alpha})^\top)^{-1}(\boldsymbol{\lambda}_1^t + \tilde{\mathbf{R}}^\top \boldsymbol{\lambda}_2^t - \tilde{\mathbf{D}}^\top \boldsymbol{\lambda}_3^t - (\boldsymbol{\Delta}\tilde{\boldsymbol{\alpha}})\boldsymbol{\lambda}_7^t + \rho \tilde{\mathbf{u}}^{t+1} - \rho \tilde{\mathbf{R}}^\top \omega^{t+1} + \rho \tilde{\mathbf{D}}^\top \mathbf{q} + \rho(\boldsymbol{\Delta}\tilde{\boldsymbol{\alpha}})(\boldsymbol{\alpha} \odot \boldsymbol{\gamma}) + \rho(\boldsymbol{\Delta}\tilde{\boldsymbol{\alpha}})(\tilde{\mathbf{I}}\tilde{\mathbf{c}}^{t+1}))$
14: **for** $i = 1, \dots, n$ **do**
15: $\mathbf{W}_i^{t+1} = \frac{1}{\rho}(\mathbf{1}\mathbf{1}^\top + \mathbf{I})^{-1}(\rho \mathbf{1}\mathbf{1}^\top + \rho \mathbf{S}_i^{t+1} - \mathbf{1}\boldsymbol{\lambda}_{4,i}^{t\top} + \boldsymbol{\Lambda}_{5,i}^t)$
16: $\mathbf{H}_i^{t+1} = \frac{1}{\rho}(\boldsymbol{\delta}\boldsymbol{\delta}^\top + \mathbf{I})^{-1}(-\boldsymbol{\delta}\boldsymbol{\lambda}_{6,i}^{t\top} + \boldsymbol{\Lambda}_{8,i}^t - \rho \boldsymbol{\delta} \beta_i^{t+1\top} + \rho \boldsymbol{\delta}(\mathbf{b}_i \odot \mathbf{B}_i \boldsymbol{\eta})^\top + \rho \mathbf{S}_i^{t+1})$
17: $\mathbf{Z}_i^{t+1} = \mathbb{1}(\rho > \tilde{\lambda})\Pi\left(\left(\frac{1}{\rho - \tilde{\lambda}}\right)(\rho \mathbf{S}_i^{t+1} + \boldsymbol{\Lambda}_{10}^t - \frac{\tilde{\lambda}}{2})\right)_{[0,1]} + \mathbb{1}(\rho < \tilde{\lambda})\Pi\left(\left(\frac{1}{\rho - \tilde{\lambda}}\right)(\rho \mathbf{S}_i^{t+1} + \boldsymbol{\Lambda}_{10}^t - \frac{\tilde{\lambda}}{2})\right)_{\{0,1\}}$
18: **end for**
19: **for** $i = 1, \dots, n$ **do**
20: $\boldsymbol{\lambda}_{1,i}^{t+1} = \boldsymbol{\lambda}_{1,i}^t + \rho(\mathbf{S}_i^{t+1}\mathbf{1} - \mathbf{u}_i^{t+1})$
21: $\boldsymbol{\lambda}_{3,i}^{t+1} = \boldsymbol{\lambda}_{3,i}^t + \rho(\mathbf{D}\mathbf{u}_i^{t+1} - \mathbf{q}_i)$
22: $\boldsymbol{\lambda}_{4,i}^{t+1} = \boldsymbol{\lambda}_{4,i}^t + \rho(\mathbf{W}_i^{t+1\top}\mathbf{1} - \mathbf{1})$
23: $\boldsymbol{\Lambda}_{5,i}^{t+1} = \boldsymbol{\Lambda}_{5,i}^t + \rho(\mathbf{S}_i^{t+1} - \mathbf{W}_i^{t+1})$
24: $\boldsymbol{\lambda}_{6,i}^{t+1} = \boldsymbol{\lambda}_{6,i}^t + \rho(\mathbf{H}_i^{t+1\top}\boldsymbol{\delta} + \beta_i^{t+1} - \mathbf{b}_i \odot \mathbf{B}_i \boldsymbol{\eta})$
25: $\boldsymbol{\Lambda}_{8,i}^{t+1} = \boldsymbol{\Lambda}_{8,i}^t + \rho(\mathbf{S}_i^{t+1} - \mathbf{W}_i^{t+1})$
26: $\boldsymbol{\Lambda}_{10,i}^{t+1} = \boldsymbol{\Lambda}_{10,i}^t + \rho(\mathbf{S}_i^{t+1} - \mathbf{Z}_i^{t+1})$
27: **end for**
28: $\boldsymbol{\lambda}_2^{t+1} = \boldsymbol{\lambda}_2^t + \rho(\boldsymbol{\omega}^{t+1} - \mathbf{R}(\sum_{i=1}^n \mathbf{u}_i^{t+1}))$
29: $\boldsymbol{\lambda}_7^{t+1} = \boldsymbol{\lambda}_7^t + \rho(\boldsymbol{\alpha} \odot (\boldsymbol{\Delta}^\top \mathbf{u}^{t+1} - \boldsymbol{\delta}) - \tilde{\mathbf{I}}\tilde{\mathbf{c}}^{t+1})$
30: $\lambda_9^{t+1} = \lambda_9^t + \rho(\tilde{\mathbf{c}}^{t+1\top}\tilde{\mathbf{1}} + \tilde{\boldsymbol{\beta}}^{t+1} - \Omega)$
31: **end for**
32: **Return:** $\mathbf{S}_i^{\tilde{T}}, \forall i = 1, \dots, n$

of the platform is that we assume the assigned routes will be followed. With autonomous vehicles, it will be easier to enforce the assigned routes. Moreover, delivery companies can enforce their drivers to follow specific routes. The ride-hailing companies can ensure compliance by incentivizing passengers/drivers who accept routes (by paying them). Another concern is protecting the privacy of individuals' data because of legal, ethical, and practical constraints. To address this concern, we can adopt approaches similar to those used in previous incentivization projects with real-world implementations [37–40]. We can also examine the scalability of our incentivization platform from various angles. Our modular design allows for the use of various prediction models, such as traffic prediction and OD estimation, tailored to different scenarios. Moreover, organizations with access to scalable real-time traffic prediction software can provide ETA predictions. The platform also offers flexibility in utilizing different VOTs for organizations with diverse operational natures.

References

[1] "INRIX 2019 global traffic scorecard," https://inrix.com/scorecard/, accessed: 2023-11-12.

[2] "Congestion mitigation and air quality improvement program: Assessing 10 years of experience," http://www.trb.org/main/blurbs/160904.aspx.

[3] Health Effects Institute. Panel on the Health Effects of Traffic-Related Air Pollution, Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects. Health Effects Institute, 2010, no. 17.

[4] K. Zhang and S. Batterman, "Air pollution and health risks due to vehicle traffic," Science of the Total Environment, vol. 450, pp. 307–316, 2013.

[5] A. C. Pigou, "The economics of welfare Macmillan and Co," London, United Kingdom, 1920.

[6] F. H. Knight, "Some fallacies in the interpretation of social cost," The Quarterly Journal of Economics, vol. 38, no. 4, pp. 582–606, 1924.

[7] E. Verhoef, M. C. Bliemer, L. Steg, and B. van Wee, Eds., Pricing in Road Transport, ser. Books. Edward Elgar Publishing, July 2008, no. 4192. [Online]. Available: https://ideas.repec.org/b/elg/eebook/4192.html

[8] D. H. Van Amelsfort, "Behavioural responses and network effects of time-varying road pricing," 2009.

[9] S. Bouchelaghem and M. Omar, "Reliable and secure distributed smart road pricing system for smart cities," IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 5, pp. 1592–1603, 2018.

[10] G. Zhang, X. Ma, and Y. Wang, "Self-adaptive tolling strategy for enhanced high-occupancy toll lane operations," IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 1, pp. 306–317, 2013.

[11] P. Kachroo, S. Gupta, S. Agarwal, and K. Ozbay, "Optimal control for congestion pricing: Theory, simulation, and evaluation," IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 5, pp. 1234–1240, 2016.

[12] F. Farokhi and K. H. Johansson, "A study of truck platooning incentives using a congestion game," IEEE Transactions on intelligent transportation systems, vol. 16, no. 2, pp. 581–595, 2014.

[13] N. Groot, B. De Schutter, and H. Hellendoorn, "Toward system-optimal routing in traffic networks: A reverse Stackelberg game approach," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 1, pp. 29–40, 2014.

[14] Z. Cao, H. Guo, W. Song, K. Gao, L. Kang, X. Zhang, and Q. Wu, "Improving the performance of transportation networks: A semi-centralized pricing approach," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 10, pp. 6353–6364, 2020.

[15] N. Zheng, G. Rérat, and N. Geroliminis, "Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment," Transportation Research Part C: Emerging Technologies, vol. 62, pp. 133–148, 2016.

[16] C. F. Daganzo and L. J. Lehe, "Distance-dependent congestion pricing for downtown zones," Transportation Research Part B: Methodological, vol. 75, pp. 89–99, 2015.

[17] S. Zhang, A. M. Campbell, and J. F. Ehmke, "Impact of congestion pricing schemes on costs and emissions of commercial fleets in urban areas," Networks, vol. 73, no. 4, pp. 466–489, 2019.

[18] R. Zhong, R. Xu, A. Sumalee, S. Ou, and Z. Chen, "Pricing environmental externality in traffic networks mixed with fuel vehicles and electric vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 9, pp. 5535–5554, 2020.

[19] J. Knockaert, Y.-Y. Tseng, E. T. Verhoef, and J. Rouwendal, "The Spitsmijden experiment: A reward to battle congestion," Transport Policy, vol. 24, pp. 260–272, 2012.

[20] D. Levinson, "Equity effects of road pricing: A review," Transport Reviews, vol. 30, no. 1, pp. 33–57, 2010.

[21] K. Martens, A. Golub, and G. Robinson, "A justice-theoretic approach to the distribution of transportation benefits: Implications for transportation planning practice in the United States," Transportation Research Part A: Policy and Practice, vol. 46, no. 4, pp. 684–695, 2012.

[22] C. Raux and S. Souche, "The acceptability of urban road pricing: A theoretical analysis applied to experience in lyon," Journal of Transport Economics and Policy (JTEP), vol. 38, no. 2, pp. 191–215, 2004.

[23] P. Ieromonachou, S. Potter, and J. P. Warren, "Evaluation of the implementation process of urban road pricing schemes in the United Kingdom and Italy," European Transport, vol. 32, pp. 49–68, 2006.

[24] D. A. Hensher and M. C. Bliemer, "What type of road pricing scheme might appeal to politicians? viewpoints on the challenge in gaining the citizen and public servant vote by staging reform," Transportation Research Part A: Policy and Practice, vol. 61, pp. 227–237, 2014.

[25] Z. Gu, Z. Liu, Q. Cheng, and M. Saberi, "Congestion pricing practices and public acceptance: A review of evidence," Case Studies on Transport Policy, vol. 6, no. 1, pp. 94–101, 2018.

[26] E. Verhoef, P. Nijkamp, and P. Rietveld, "Tradeable

permits: their potential in the regulation of road transport externalities," Environment and Planning B: Planning and Design, vol. 24, no. 4, pp. 527–548, 1997.

[27] G. Wang, Z. Gao, M. Xu, and H. Sun, "Models and a relaxation algorithm for continuous network design problem with a tradable credit scheme and equity constraints," Computers & Operations Research, vol. 41, pp. 252–261, 2014.

[28] W. Fan and X. Jiang, "Tradable mobility permits in roadway capacity allocation: Review and appraisal," Transport Policy, vol. 30, pp. 132–142, 2013.

[29] J. Thøgersen and B. Møller, "Breaking car use habits: The effectiveness of a free one-month travelcard," Transportation, vol. 35, no. 3, pp. 329–345, 2008.

[30] T. Tsekeris and S. Voß, "Design and evaluation of road pricing: state-of-the-art and methodological advances," NETNOMICS: Economic Research and Electronic Networking, vol. 10, no. 1, pp. 5–52, 2009.

[31] H. Fukui, "An empirical analysis of airport slot trading in the United States," Transportation Research Part B: Methodological, vol. 44, no. 3, pp. 330–357, 2010.

[32] N. Dogterom, D. Ettema, and M. Dijst, "Tradable credits for managing car travel: a review of empirical research and relevant behavioural approaches," Transport Reviews, vol. 37, no. 3, pp. 322–343, 2017.

[33] C. L. Azevedo, R. Seshadri, S. Gao, B. Atasoy, A. P. Akkinepally, E. Christofa, F. Zhao, J. Trancik, and M. Ben-Akiva, "Tripod: sustainable travel incentives with prediction, optimization, and personalization," in Transportation Research Board 97th Annual Meeting, 2018.

[34] J. W. Brehm, "A theory of psychological reactance." 1966.

[35] D. M. Kreps, "Intrinsic motivation and extrinsic incentives," The American Economic Review, vol. 87, no. 2, pp. 359–364, 1997.

[36] K. C. Berridge, "Reward learning: reinforcement, incentives, and expectations." 2001.

[37] D. Merugu, B. S. Prabhakar, and N. Rama, "An incentive mechanism for decongesting the roads: A pilot program in bangalore," in Proc. of ACM NetEcon Workshop. Citeseer, 2009.

[38] J. S. Yue, C. V. Mandayam, D. Merugu, H. K. Abadi, and B. Prabhakar, "Reducing road congestion through incentives: a case study," 2015.

[39] M. Bliemer, M. Dicke-Ogenia, and D. Ettema, "Rewarding for avoiding the peak period: A synthesis of three studies in the Netherlands," in European Transport Conference 2009. Citeseer, 2009.

[40] X. Hu, Y.-C. Chiu, and L. Zhu, "Behavior insights for an incentive-based active demand management platform," International Journal of Transportation Science and Technology, vol. 4, no. 2, pp. 119–133, 2015.

[41] J. Wu, Y. Tian, and J. Sun, "Managing ridesharing with incentives in a bottleneck model," Research in Transportation Economics, vol. 101, p. 101349, 2023.

[42] J. Wu, Y. Tian, J. Sun, H. M. Zhang, and Y. Wang, "Public or private? optimal organization for incentive-based travel demand management," Transportation Research Part E: Logistics and Transportation Review, vol. 174, p. 103137, 2023.

[43] V. Kumar, C. R. Bhat, R. M. Pendyala, D. You, E. Ben-Elia, and D. Ettema, "Impacts of incentive-based intervention on peak period traffic: experience from the Netherlands," Transportation Research Record, vol. 2543, no. 1, pp. 166–175, 2016.

[44] A. He, "People continue to rely on maps and navigational apps," https://www.emarketer.com/content/people-continue-to-rely-on-maps-and-navigational-apps-emarketer-forecasts-show, accessed: 2023-11-12.

[45] R. Panko, "The popularity of google maps: Trends in navigation apps in 2018," https://themanifest.com/app-development/trends-navigation-apps, accessed: 2023-11-12.

[46] "Grocery continues to power growth for doordash," https://www.grocerydive.com/news/grocery-continues-to-power-growth-for-doordash/643120/, accessed: 2023-11-12.

[47] "Uber announces results for fourth quarter and full year 2022," https://investor.uber.com/news-events/news/press-release-details/2023/Uber-Announces-Results-for-Fourth-Quarter-and-Full-Year-2022/default.aspx, accessed: 2023-11-12.

[48] A. Ghafelebashi, M. Razaviyayn, and M. Dessouky, "Congestion reduction via personalized incentives," Transportation Research Part C: Emerging Technologies, vol. 152, p. 104153, 2023.

[49] ——, "Incentive systems for fleets of new mobility services," arXiv preprint arXiv:2312.02341, 2023.

[50] U. S. B. of Public Roads, Traffic assignment manual for application with a large, high speed computer. US Department of Commerce, Bureau of Public Roads, Office of Planning, Urban Planning Division, 1964, vol. 37.

[51] W. Ma and Z. S. Qian, "Estimating multi-year 24/7 origin-destination demand using high-granular multi-source traffic data," Transportation Research Part C: Emerging Technologies, vol. 96, pp. 96–121, 2018.

[52] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[53] ——, "Graph implementations for nonsmooth convex programs," in Recent Advances in Learning and Control, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[54] S. Boyd, N. Parikh, and E. Chu, Distributed optimization and statistical learning via the alternating direction method of multipliers. Now Publishers Inc, 2011.

[55] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," Computers & Mathematics with Applications, vol. 2, no. 1, pp. 17–40, 1976.

[56] R. Glowinski and A. Marroco, "Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires," ESAIM: Mathematical Modelling

and Numerical Analysis-Modélisation Mathématique et Analyse Numérique, vol. 9, no. R2, pp. 41–76, 1975.

[57] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," SIAM Journal on Optimization, vol. 26, no. 1, pp. 337–364, 2016.

[58] B. Barazandeh, A. Ghafelebashi, M. Razaviyayn, and R. Sriharsha, "Efficient algorithms for estimating the parameters of mixed linear regression models," arXiv preprint arXiv:2105.05953, 2021.

[59] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50–60, 2020.

[60] A. Lowy, A. Ghafelebashi, and M. Razaviyayn, "Private non-convex federated learning without a trusted server," in International Conference on Artificial Intelligence and Statistics. PMLR, 2023, pp. 5749–5786.

[61] C. Anastasiou, J. Lin, C. He, Y.-Y. Chiang, and C. Shahabi, "Admsv2: A modern architecture for transportation data management and analysis," in Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Advances on Resilient and Intelligent Cities, 2019, pp. 25–28.

[62] H. J. Van Zuylen and L. G. Willumsen, "The most likely trip matrix estimated from traffic counts," Transportation Research Part B: Methodological, vol. 14, no. 3, pp. 281–293, 1980.

[63] E. Cascetta, "Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator," Transportation Research Part B: Methodological, vol. 18, no. 4-5, pp. 289–299, 1984.

[64] M. G. Bell, "The real time estimation of origin-destination flows in the presence of platoon dispersion," Transportation Research Part B: Methodological, vol. 25, no. 2-3, pp. 115–125, 1991.

[65] S. Bera and K. Rao, "Estimation of origin-destination matrix from traffic counts: the state of the art," European Transport\Trasporti Europei, 2011.

[66] P. Krishnakumari, H. van Lint, T. Djukic, and O. Cats, "A data driven method for od matrix estimation," Transportation Research Part C: Emerging Technologies, 2019.

[67] S. Carrese, E. Cipriani, L. Mannini, and M. Nigro, "Dynamic demand estimation and prediction for traffic urban networks adopting new data sources," Transportation Research Part C: Emerging Technologies, vol. 81, pp. 83–98, 2017.

[68] M. Nigro, E. Cipriani, and A. del Giudice, "Exploiting floating car data for time-dependent origin–destination matrices estimation," Journal of Intelligent Transportation Systems, vol. 22, no. 2, pp. 159–174, 2018.

[69] J. Kim, F. Kurauchi, N. Uno, T. Hagihara, and T. Daito, "Using electronic toll collection data to understand traffic demand," Journal of Intelligent Transportation Systems, vol. 18, no. 2, pp. 190–203, 2014.

[70] J. C. Castillo, "Who benefits from surge pricing?" Available at SSRN 3245533, 2020.

Ali Ghafelebashi received his B.Sc. degree in Industrial and Systems Engineering from Amirkabir University of Technology in 2018. He obtained his M.Sc. degree in Computer Science from the University of Southern California. He is currently pursuing a Ph.D. degree in Industrial and Systems Engineering at the University of Southern California. He also serves as a Research Assistant with the Optimization for Data-Driven Science (ODDS) Research Group, University of Southern California, advised by Professor Meisam Razaviyayn. Ali Ghafelebashi is the recipient of the 2023 ITS California (ITSCA) and California Transportation Foundation (CTF) Scholarships and the 2023 University of Southern California Viterbi's 3MT Director's Award. His research interests are at the intersection of large-scale optimization, transportation, and machine learning.

Meisam Razaviyayn is an associate professor of Industrial and Systems Engineering, Computer Science, and Electrical Engineering at the University of Southern California. He is also the associate director of the USC- Meta Center for Research and Education in AI and Learning. Prior to joining USC, he was a postdoctoral research fellow in the Department of Electrical Engineering at Stanford University. He received his PhD in Electrical Engineering with a minor in Computer Science at the University of Minnesota. He obtained his M.Sc. degree in Mathematics from the University of Minnesota. Meisam Razaviyayn is the recipient of the 2022 NSF CAREER Award, the 2022 Northrop Grumman Excellence in Teaching Award, the 2021 AFOSR Young Investigator Award, the 2021 3M Nontenured Faculty Award, 2020 ICCM Best Paper Award in Mathematics, IEEE Data Science Workshop Best Paper Award in 2019, the Signal Processing Society Young Author Best Paper Award in 2014, and the finalist for Best Paper Prize for Young Researcher in Continuous Optimization in 2013 and 2016. He is also the silver medalist of Iran's National Mathematics Olympiad. His research interests include the design and the study of the fundamental aspects of optimization algorithms that arise in the modern data science era.

Maged Dessouky received the B.S. and M.S. degrees from Purdue University and the Ph.D. degree in industrial engineering from the University of California, Berkeley. He is currently a Professor and Chair in Industrial and Systems Engineering at the University of Southern California. He received the 2007 Transportation Science and Logistics Best Paper Prize. He is an Area/Associate Editor of Transportation Research Part B: Methodological, IIE Transactions, and Computers and Industrial Engineering, is on the Editorial Board of Transportation Research Part E: Logistics and Transportation Review, and has previously served as an Area Editor for the ACM Transactions of Modeling and Computer Simulation and as an Associate Editor for the IEEE Transactions on Intelligent Transportation Systems. He is a fellow of INFORMS and IISE.