Human-in-the-loop Learning for Dynamic Congestion Games

Hongbo Li, *Student Member, IEEE* and Lingjie Duan, *Senior Member, IEEE*

***Abstract*—Today mobile users learn and share their traffic observations via crowdsourcing platforms (e.g., Google Maps and Waze). Yet such platforms simply cater to selfish users' myopic interests to recommend the shortest path, and do not encourage enough users to travel and learn other paths for future others. Prior studies focus on one-shot congestion games without considering users' information learning, while our work studies how users learn and alter traffic conditions on stochastic paths in a human-in-the-loop manner. In a typical parallel routing network with one deterministic path and multiple stochastic paths, our analysis shows that the myopic routing policy (used by Google Maps and Waze) leads to severe under-exploration of stochastic paths. This results in a price of anarchy (PoA) greater than 2, as compared to the socially optimal policy achieved through optimal exploration-exploitation tradeoff in minimizing the long-term social cost. Besides, the myopic policy fails to ensure the correct learning convergence about users' traffic hazard beliefs. To address this, we focus on informational (non-monetary) mechanisms as they are easier to implement than pricing. We first show that existing information-hiding mechanisms and deterministic path-recommendation mechanisms in Bayesian persuasion literature do not work with even PoA $= \infty$. Accordingly, we propose a new combined hiding and probabilistic recommendation (CHAR) mechanism to hide all information from a selected user group and provide state-dependent probabilistic recommendations to the other user group. Our CHAR mechanism successfully ensures PoA less than $\frac{5}{4}$, which cannot be further reduced by any other informational (non-monetary) mechanism. Besides the parallel network, we further extend our analysis and CHAR mechanism to more general linear path graphs with multiple intermediate nodes, and we prove that the PoA results remain unchanged. Additionally, we carry out experiments with real-world datasets to further extend our routing graphs and verify the close-to-optimal performance of our CHAR mechanism.**

***Index Terms*—human-in-the-loop learning, dynamic congestion games, price of anarchy, mechanism design**

I. Introduction

In today's traffic networks, a random number of users arrive sequentially to decide their routing paths (e.g., the shortest path) based on real-time traffic conditions [2]. To assist new user arrivals in their decision-making process, emerging crowdsourcing applications (e.g., Google Maps and Waze) serve as important platforms to leverage users for information learning and sharing ([3], [4]). These platforms aggregate the latest traffic information of stochastic paths from previous users who have recently traveled there and reported their observations, then share this information with new users. However, these platforms always myopically recommend the currently shortest path, and do not encourage enough selfish users to travel and learn other longer paths of varying traffic conditions for future others. Prior congestion game literature assumes a static scenario of one-shot routing by a group of users, without considering their long-term information learning in a long-term dynamic scenario (e.g., [5], [6]).

To efficiently learn and share the time-varying information, multi-armed bandit (MAB) problems are developed to study the optimal exploration-exploitation among stochastic arms/paths (e.g., [7]–[11]). For example, [10] utilizes MAB techniques to predict congestion in a fast-varying transportation environment. [11] applies the MAB exploration-exploitation strategy to balance the recruitment of previously well-behaved users or new uncertain users for completing crowdsourcing tasks. Recently, some studies have extended traditional MAB problems to distributed scenarios with multiple user arrivals at the same time (e.g., [12]–[16]). For instance, [14] study how users make their own optimal decisions at each time slot based on the private arm information that they learned beforehand. [13] explores cooperative information exchange among all users to facilitate local model learning. Subsequently, both [15] and [16] examine the performance under partially connected communication graphs, where users communicate only with their neighbors to make locally optimal decisions. In general, these MAB solutions assume users' cooperation to align with the social planner, without possible selfish deviations to improve their own benefits.

As selfish users may not listen to the social planner, some recent congestion game works focus on Bayesian persuasion and information design to incentivize users' exploration of different paths [18]–[22]. For example, [20] assumes that the social planner has the complete system information and properly discloses system information to regulate users and control efficiency loss. Without perfect feedback, [21] estimates users' expected utilities of traveling each path based on all possible cases to improve the robustness of Bayesian persuasion. Similarly, [22] allows the system to learn stochastic travel latency based on users' travel data while only regulating the immediate congestion in the current time slot. In summary, all these works assume that users' routing decisions do not internally alter the long-term traffic congestion condition for future users, and

This work was supported in part by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 Grant with Award no. MOE-T2EP20121-0001; in part by SUTD Kickstarter Initiative (SKI) Grant with no. SKI 2021_04_07; and in part by the Joint SMU-SUTD Grant with no. 22-LKCSB-SMU-053. Part of this work was published in IEEE ISIT 2024 Conference [1].

Hongbo Li and Lingjie Duan are with the Pillar of Engineering Systems and Design, Singapore University of Technology and Design, Singapore 487372 (e-mail: hongbo_li@mymail.sutd.edu.sg; lingjie_duan@sutd.edu.sg).

only focus on exogenous information to learn dynamically.

It is practical to consider endogenous information variation in dynamic congestion games, where more users choosing the same path not only improve learning accuracy there ([23]), but also produce more congestion for followers ([24], [25]). For this new problem, we need to overcome two technical issues. The first is how to *dynamically allocate users to reach the best trade-off between learning accuracy and resultant congestion costs of stochastic paths*. To minimize the long-run social travel cost, it is critical to dynamically balance multiple user arrivals' positive information learning effect and negative congestion effect to avoid both under- and over-exploration of stochastic paths.

Even provided with the socially optimal policy, we still need to ensure selfish users follow it. Thus, the second issue is how to *design an informational mechanism to properly regulate users' myopic routing*. In practice, informational mechanisms (e.g., Bayesian persuasion [18]–[20]) are non-monetary and easier to implement compared to pricing or side-payment mechanisms (e.g., [26]–[28]). There are two commonly used informational mechanisms to optimize long-run information learning in congestion games: information hiding mechanism [29]–[31] and deterministic path-recommendation [32]–[34]. For example, [29] analyzes how users infer the expected travel cost of each path to decide routing if the system hides all traffic information. [34] assumes a publicly known process for varying each path's traffic condition, and selectively provides deterministic path recommendations to regulate users' myopic routing. We will show later that both mechanisms do not work in improving our system performance, inspiring our new mechanism design.

Our main contributions and the key novelty of this paper are summarized as follows.

- *Human-in-the-loop learning for dynamic congestion games:* Prior studies focus on one-shot congestion games without information learning, while we study how users learn and alter traffic conditions on stochastic paths in a human-in-the-loop manner. When such congestion games meet human-in-the-loop learning, more users' routing on stochastic paths generates both positive learning benefits and negative congestion effects there for followers. We use users' positive learning effect to negate negative congestion, which fundamentally extends traditional one-shot congestion games (e.g., [20], [22] and [33]).
- *Policies comparison via PoA analysis:* To minimize the social cost, we formulate the optimization problems for both myopic and socially optimal policies as Markov decision processes (MDP). Our analysis shows that the myopic routing policy (used by Google Maps and Waze) leads to severe under-exploration of stochastic paths. This results in a price of anarchy (PoA) greater than 2, as compared to the socially optimal policy achieved through optimal exploration-exploitation tradeoff in minimizing the long-term social cost.
- *Learning convergence and new mechanism design:* We first prove that the socially optimal policy ensures correct convergence of users' traffic hazard beliefs on stochastic paths, while the myopic policy cannot. Then we show that existing information-hiding and deterministic-recommendation mechanisms in Bayesian persuasion literature make PoA $= \infty$. Accordingly, we propose a combined hiding and probabilistic recommendation (CHAR) mechanism to hide all information from a selected user group and provide state-dependent probabilistic recommendations to the other user group. Our CHAR mechanism successfully ensures PoA less than $\frac{5}{4}$, which cannot be further reduced by any other informational mechanism.
- *Extensions to linear path graph:* Besides the basic parallel transportation network between two nodes, we further extend our analysis and CHAR mechanism to encompass any linear path graph with multiple intermediate nodes. In this extended model, myopic users also consider their future travel costs when making current routing decisions, which adds complexity to their decision analysis. Nonetheless, we successfully prove that the PoA caused by users' myopic routing policy remains greater than 2, and our CHAR mechanism continues to effectively reduce PoA to less than $\frac{5}{4}$. Finally, we conduct experiments using real-world datasets to show the close-to-optimal average-case performance of our CHAR mechanism.

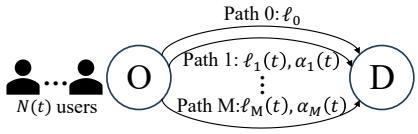

Fig. 1. Dynamic congestion model: within each time slot $t \in \{0, 1, \cdots\}$, a random number $N(t)$ of users arrive at origin O to decide among safe path 0 and any stochastic path $i \in \mathcal{M} := \{1, \cdots, M\}$ to travel to destination D.

The rest of the paper is organized as follows. In Section II, we formulate the dynamic congestion model and the human-in-the-loop learning model for the system. Then we formulate the optimization problems for both myopic policy and socially optimal policy in Section III. After that, in Section IV we analytically compare the two policies via PoA analysis. Based on the analysis, we propose our CHAR mechanism in Section V. Subsequently, we extend the system model and analysis to a general linear path graph in Section VI and verify the performance of our CHAR mechanism in Section VII. Finally, Section VIII concludes the paper. For ease of reading, we summarize all the key notations in Table I.

II. System Model

In this section, we first introduce the dynamic congestion model for a typical parallel multi-path network between two nodes O and D as in existing congestion game literature (e.g., [29], [33] and [34]). Then we introduce the human-in-the-loop learning model for the crowdsourcing platform. Note that we have extended our congestion model and key results to any general linear path graphs later in Section VI. As shown in Fig. 1, we consider an infinite discrete time horizon $t \in \{0, 1, \cdots\}$ to model the dynamic congestion game. Within each time slot t, a stochastic number $N(t) \in [N_{min}, N_{max}]$ of atomic users, where $\mathbb{E}[N(t)] = \bar{N}$, arrive at origin O to decide their paths among $M + 1$ paths to travel to destination D.

TABLE I
KEY NOTATIONS AND THEIR MEANINGS IN THE PAPER

Notation	Meaning
$N(t)$	The number of user arrivals at time t.
N_{min}, N_{max}	The lower and upper bounds of $N(t)$.
$\bar{N}$	The expectation value of $N(t)$.
$\mathcal{M}$	The set of stochastic paths.
ℓ_0	The fixed travel latency of path 0.
$\ell_i(t)$	The travel latency of path i at time t.
$n_i(t)$	The number of users traveling path i at time t.
$\alpha_i(t)$	The correlation coefficient of stochastic path i.
α_H, α_L	The high and low hazard states on stochastic paths.
$\bar{x}$	The long-run expected probability of $\alpha_i(t) = \alpha_H$.
$\mathbb{P}(\bar{x})$	The distribution of $\bar{x}$.
$y_i(t)$	Users' observation summary of path i at time t.
$q_H(\cdot), q_L(\cdot)$	The probabilities for $n_i(t)$ users to report $y_i(t) = 1$ under α_H and α_L, respectively.
$x_i(t), x_i'(t)$	The prior and posterior hazard beliefs of stochastic path i at time t.
$\mathbf{L}(t)$	The expected travel latency set of all $M+1$ paths.
$\mathbf{x}(t)$	The hazard belief set of M stochastic paths.
$c_i(\cdot)$	Each user's travel cost on path i at time t.
$V(\cdot)$	The variance cost at time t.
$C(\cdot)$	The long-term cost function.
ρ	Discount factor.
$\pi(t)$	The private path recommendation for each user under our CHAR mechanism at time t.

A. Dynamic Congestion Model

Following the traditional routing game literature ([18], [34], [35]), we model a safe path 0 with a fixed travel latency, which is denoted by ℓ_0. While the other paths are stochastic with time-varying travel latencies, denoted by $\ell_i(t)$ at time $t \in \{1, 2, \cdots\}$, where $i \in \mathcal{M} := \{1, \cdots, M\}$. According to prior research on travel delay estimation (e.g., [36], [37]), the next $\ell_i(t+1)$ at $t+1$ is correlated with both current $\ell_i(t)$ and the number of atomic users traveling this path, which is denoted by $n_i(t)$. For example, more users traveling on the same path results in slower driving speeds and longer queueing times at traffic lights, thereby increasing the latency for later users. Define this general correlation function as:

$$\ell_i(t+1) = f\big(\ell_i(t), n_i(t), \alpha_i(t)\big), \tag{1}$$

where $\alpha_i(t)$ is the correlation coefficient, measuring the left-over flow to be served over time. Note that $f(\cdot)$ is defined as a general increasing function in $\ell_i(t), n_i(t)$ and $\alpha_i(t)$.

As in existing congestion game literature ([31], [33], [34]), we model $\alpha_i(t)$ to follow a memoryless stochastic process (not necessarily a time-invariant Markov chain) to alter as below:

$$\alpha_i(t) = \begin{cases} \alpha_H, \text{if high hazard (bad) condition on path } i \text{ at } t, \\ \alpha_L, \text{if low hazard (good) condition on path } i \text{ at } t, \end{cases}$$

where $\alpha_H \in [1, +\infty)$ and $\alpha_L \in [0, 1)$. Unlike these works assuming that the social planner knows the probability distribution of $\alpha_i(t)$ beforehand, we relax this assumption to keep it unknown in our work. For ease of exposition, we only consider two states here, while our later analysis and mechanism design can be extended to accommodate multiple states of $\alpha_i(t)$.

We denote the expected steady state of correlation coefficient $\alpha_i(t)$ in the long run by

$$\mathbb{E}[\alpha_i(t)|\bar{x}] = \bar{x} \cdot \alpha_H + (1 - \bar{x}) \cdot \alpha_L, \tag{2}$$

where $\bar{x}$ is the long-run expected probability of $\alpha_i(t) = \alpha_H$ yet the exact value of $\bar{x}$ is unknown to users. They can only know $\bar{x}$ to satisfy a distribution $\mathbb{P}(\bar{x})$. Here we assume $\mathbb{P}(\bar{x})$ is not extreme, such that users may consider traveling on any stochastic path i under $\bar{x} \sim \mathbb{P}(\bar{x})$. To accurately estimate $\mathbb{E}[\alpha_i(t)|\bar{x}]$, the system expects users to learn the real steady state $\bar{x}$ by observing real-time traffic conditions.

B. Human-in-the-loop Learning Model

When traveling on stochastic path i, users cannot observe $\alpha_i(t)$ directly but a traffic hazard event (e.g., 'black ice' segments or jamming). At the same time, [38] shows that each user has a noisy observation of the hazard. Hence, our system uses a majority vote to fuse all their observation reports of stochastic path i into a hazard summary $y_i(t) \in \{0, 1, \emptyset\}$ during time t. Specifically,

- $y_i(t) = 1$ tells that most of the current $n_i(t)$ users observe a traffic hazard on stochastic path i at time t.
- $y_i(t) = 0$ tells that most of the current $n_i(t)$ users observe no traffic hazard on stochastic path i at time t.
- $y_i(t) = \emptyset$ tells that there is no user traveling on path i (i.e., $n_i(t) = 0$), without any new observation.

However, the fused summary $y_i(t)$ by $n_i(t)$ users can still be inaccurate. Given $n_i(t)$, we define the group probabilities under $\alpha_i(t)_H$ and $\alpha_i(t) = \alpha_L$ for observing a hazard below:

$$\begin{aligned} q_H(n_i(t)) &= \mathbf{Pr}\big(y_i(t) = 1 | \alpha_i(t) = \alpha_H, n_i(t)\big), \\ q_L(n_i(t)) &= \mathbf{Pr}\big(y_i(t) = 1 | \alpha_i(t) = \alpha_L, n_i(t)\big). \end{aligned} \tag{3}$$

According to [23], $q_H(n_i(t)) \in [0, 1]$ increases with $n_i(t)$, because more users help spot hazard $\alpha_i(t) = \alpha_H$. Similarly, $q_L(n_i(t)) \in [0, 1]$ decreases with $n_i(t)$ under $\alpha_i(t) = \alpha_L$.

We summarize the current routing decisions and observations of all paths as vectors $\mathbf{n}(t) = \{n_i(t) | i \in \{0, \cdots, M\}\}$ and $\mathbf{y}(t) = \{y_i(t) | i \in \mathcal{M}\}$, respectively. To guide trip advisory at time t, the crowdsourcing platform needs to memorize all the historical data of $(\mathbf{n}(1), \cdots, \mathbf{n}(t-1))$ and $(\mathbf{y}(1), \cdots, \mathbf{y}(t-1))$, which keep growing over time. Following the memoryless property [32], we use Bayesian inference to equivalently summarize these data into a prior hazard belief $x_i(t) \in [0, 1]$ for any stochastic path i:

$$x_i(t) = \mathbf{Pr}\big(\alpha_i(t) = \alpha_H | x_i(t-1), y_i(t-1), n_i(t-1)\big), \tag{4}$$

which indicates the probability of $\alpha_i(t) = \alpha_H$ at time t.

Next, we explain how our human-in-the-loop learning model works in the time horizon.

- At the beginning of time t, the crowdsourcing platform publishes latest latency $\mathbb{E}[\ell_i(t)|x_i(t-1), y_i(t-1)]$ and prior hazard belief $x_i(t)$ in (4) on any stochastic path i.
- During time t, $N(t)$ users arrive to decide $n_i(t)$ on each stochastic path i and travel there to return their

observation summary $y_i(t)$. Based on $y_i(t)$, the platform next updates prior belief $x_i(t)$ to a posterior belief $x_i'(t) = \mathbf{Pr}(\alpha_i(t) = \alpha_H | x_i(t), y_i(t), n_i(t))$. For example, if $y_i(t) = 1$, we have

$$\begin{aligned} x_i'(t) &= \Pr(\alpha_i(t) = \alpha_H | x_i(t), n_i(t), y_i(t) = 1) \\ &= \frac{\Pr(y_i(t) = 1 | \alpha_i(t) = \alpha_H) \Pr(\alpha_i(t) = \alpha_H)}{\sum_{j \in \{\alpha_H, \alpha_L\}} \Pr(y_i(t) = 1 | \alpha_i(t) = j) \Pr(\alpha_i(t) = j)} \\ &= \frac{x_i(t) q_H(n_i(t))}{x_i(t) q_H(n_i(t)) + (1 - x_i(t)) q_L(n_i(t))} \end{aligned} \tag{5}$$

where the second equation is due to Bayes' Theorem and Law of total probability. Similarly, if $y_i(t) = 0$, we obtain

$$x'(t) = \frac{x(t)\big(1 - q_H(n(t))\big)}{x(t)\big(1 - q_H(n(t))\big) + (1 - x(t))\big(1 - q_L(n(t))\big)}.$$

Note that we keep $x'(t) = x(t)$ if $y_i(t) = \emptyset$, as there is no observation to update $x'(t)$.

- At the end of time t, the platform updates the expected correlation coefficient given posterior belief $x_i'(t)$:

$$\mathbb{E}[\alpha_i(t) | x_i'(t)] = x_i'(t)\alpha_H + (1 - x_i'(t))\alpha_L. \tag{6}$$

Combing (6) with (1), we obtain expected travel latency

$$\mathbb{E}[\ell_i(t+1) | x_i'(t)] = \mathbb{E}\big[f\big(\mathbb{E}[\ell_i(t) | x_i'(t-1)], n_i(t), \alpha_i(t)\big)\big] \tag{7}$$

on stochastic path i for next time slot $t + 1$.

Finally, the platform updates $x_i(t+1) = x_i'(t)$ and repeats above process in $t + 1$.

III. Problem Formulations for Myopic and Socially Optimal Policies

Based on the dynamic congestion and human-in-the-loop learning models in Section II, we next formulate the optimization problems for both myopic and socially optimal policies. The myopic policy is widely used in existing crowdsourcing platforms (e.g., Google Maps and Waze), defined as follows.

Definition 1 (Myopic Policy): Under the myopic policy, the platform recommends a path to each user to minimize their current travel cost.

A. Problem Formulation for the Myopic Policy

In this subsection, we focus on the myopic policy used by Google Maps and Waze, which post all the collected useful information to the public. First, we summarize all stochastic paths' expected latencies and hazard beliefs into vectors

$$\begin{aligned} \mathbf{L}(t) &= \{\mathbb{E}[\ell_i(t) | x_i'(t-1)] | i \in \mathcal{M}\}, \\ \mathbf{x}(t) &= \{x_i(t) | i \in \mathcal{M}\}, \end{aligned}$$

respectively. Let $n_i^{(m)}(t) \in \{0, \cdots, N(t)\}$ define the (exploration) number of users traveling on any path $i \in \{0, \cdots, M\}$ allocated by the myopic policy from $N(t)$ users, where the superscript (m) means the myopic policy. We use $\mathbf{n}^{(m)}(t)$ to summarize all the $n_i^{(m)}(t)$ under the myopic policy.

Following the typical travel cost model as in congestion game literature [18], [29], for each of the $n_0(t)$ users on safe path 0, his expected travel cost consists of travel latency ℓ_0 and congestion cost $n_0(t)$ caused by other users on this path:

$$c_0(n_0(t)) = \ell_0 + n_0(t). \tag{8}$$

For each user on stochastic path $i \in \mathcal{M}$, besides the expected travel latency $\mathbb{E}[\ell_i(t) | x_i'(t-1)]$ caused by past users and the immediate congestion cost $n_i(t)$ in traditional congestion games without learning ([18], [29]), he also faces an extra variance cost $V(n_i(t-1))$ due to former $n_i(t-1)$ users' imperfect observation summary $y_i(t-1)$ ([33], [39]). $V(n_i(t-1))$ tells how much the gap (between the $n_i(t-1)$ reports fused result and the reality) adds to the total cost by misleading decision making of future users. Then his travel cost on stochastic path i is:

$$c_i(n_i(t)) = \mathbb{E}[\ell_i(t) | x_i'(t-1)] + n_i(t) + V\big(n_i(t-1)\big), \tag{9}$$

where $V(n_i(t-1))$ is a general decreasing function of $n_i(t-1)$, as more users improve the learning accuracy on stochastic path i.

Based on each user's individual cost (8) of safe path 0 and (9) of stochastic path i, we next analyze myopic policy $n_i^{(m)}(t)$. For ease of exposition, we first consider $M = 1$ in a basic two-path network to solve and explain $n_1^{(m)}(t)$ below.

Lemma 1: Under the myopic policy, given expected travel latency $\mathbb{E}[\ell_1(t) | x_1'(t-1)]$ and hazard belief $x_1(t)$ of stochastic path 1, the exploration number is:

$$n_1^{(m)}(t) = \begin{cases} 0, & \text{if } c_1(0) \geq c_0(N(t)), \\ N(t), & \text{if } c_1(N(t)) \leq c_0(0), \\ \frac{N(t)}{2} + \frac{c_0(0) - c_1(0)}{2}, & \text{otherwise.} \end{cases} \tag{10}$$

The proof of Lemma 1 is given in Appendix A of the supplement material. If the minimum expected travel cost of stochastic path 1 is larger than the maximum cost of path 0, all the $N(t)$ users will choose path 0 with $n_1^{(m)}(t) = 0$ in the first case of (10). Otherwise, in the last two cases of (10), there is always a positive $n_1^{(m)}(t) \geq 1$ number of users traveling on stochastic path 1 to learn and update $x_1(t+1)$ and $\mathbb{E}[\ell_1(t+1) | x_1'(t)]$ there. When there are $M \geq 2$ stochastic paths, $N(t)$ user arrivals myopically decide on the exploration number $n_i(t)$ for each path i, as described in (10), in an attempt to balance the expected cost $c_i(n_i^{(m)}(t)) = c_j(n_j^{(m)}(t))$ for any path $j \neq i$ at the equilibrium. Under the constraint $\sum_{i=0}^{M} n_i^{(m)}(t) = N(t)$, this system of $M + 1$ linear equations in $M + 1$ unknowns can be solved by Gaussian elimination.

Based on the above analysis, we further examine the long-run social cost of all users since the current time t. For current $N(t)$ user arrivals, their immediate social cost under the myopic policy is

$$c(\mathbf{n}^{(m)}(t)) = \sum_{i=0}^{M} n_i^{(m)}(t) c_i(n_i^{(m)}(t)) \tag{11}$$

with individual cost $c_i(\cdot)$ in (8) or (9). Define the long-term ρ-discounted social cost function under the myopic policy to

be $C^{(m)}(\mathbf{L}(t), \mathbf{x}(t), N(t))$. Based on (11), we leverage the Markov decision process (MDP) to formulate:

$$C^{(m)}\big(\mathbf{L}(t), \mathbf{x}(t), N(t)\big) = \quad (12)$$
$$c(\mathbf{n}^{(m)}(t)) + \rho\mathbb{E}\big[C^{(m)}\big(\mathbf{L}(t+1), \mathbf{x}(t+1), N(t+1)\big)\big],$$

where the update of each travel latency $\mathbb{E}[\ell_i(t+1)|x_i'(t)]$ and hazard belief $x_i(t+1)$ are given in (5) and (7), depending on current $N(t)$ users' routing decision $n_i^{(m)}(t)$ and the possible observation summary $y_i(t) \in \{0, 1, \emptyset\}$ on path i. Here the discount factor $\rho \in (0, 1)$ is widely used to discount future travel costs to its present cost [40].

B. Problem Formulation for Socially Optimal Policy

Different from the myopic policy, the socially optimal policy minimizes the long-term social travel cost for all users. Define $n_i^*(t) \in \{0, 1, \cdots, N(t)\}$ to be the optimal exploration number, and let vector $\mathbf{n}^*(t)$ summarize all the $n_i^*(t)$. Different from the myopic policy that only minimizes each user's travel cost, the social optimum aims to well control the overall congestion cost and information learning to minimize the long-run expected social cost.

Define the long-run expected social cost function under the socially optimal policy to be $C^*(\mathbf{L}(t), \mathbf{x}(t), N(t))$. Under the socially optimal policy, the update of $\mathbf{L}(t)$ and $\mathbf{x}(t)$ also depend on the optimal exploration number set $\mathbf{n}^*(t)$. For any stochastic path i,

- If $n_i^*(t) = 0$, i.e., the platform asks no user to choose this stochastic path i, then $y_i(t) = \emptyset$ and the next hazard belief $x_i(t+1) = x_i(t)$ keeps unchanged.
- Otherwise, $x_i(t+1)$ is updated based on $y_i(t) = 1$ or 0.

After that, the expected travel latency $\mathbb{E}[\ell_i(t+1)|x_i'(t)]$ of each stochastic path i is updated according to (7).

Based on the above analysis, we formulate the long-term objective function under the socially optimal policy as:

$$C^*\big(\mathbf{L}(t), \mathbf{x}(t), N(t)\big) = \quad (13)$$
$$\min_{\mathbf{n}^*(t)} \; c(\mathbf{n}^*(t)) + \rho\mathbb{E}\big[C^*\big(\mathbf{L}(t+1), \mathbf{x}(t+1), N(t+1)\big)\big],$$

where the immediate social cost $c(\mathbf{n}^*(t))$ is similarly defined as in (11). Note that (13) is non-convex and incurs the curse of dimensionality under the infinite time horizon [41].

To derive an approximated solution $\mathbf{n}^*(t)$, we combine successive approximation [42] and exhaustive search [43] to formulate the following iterative algorithm, which iterates over Q steps. In each step $q \in \{1, \cdots, Q\}$, we update the current solution $\mathbf{n}_q(t)$ by solving the problem $C_q(\mathbf{L}(t), \mathbf{x}(t), N(t))$.

Initially, we define

$$C_0(\mathbf{L}(t), \mathbf{x}(t), N(t)) = \min_{\mathbf{n}_0(t)} \; c(\mathbf{n}_0(t)),$$

which is a linear optimization problem and can be solved by the ellipsoid method [41]. Then for any step $q \geq 1$, let

$$C_q(\mathbf{L}(t), \mathbf{x}(t), N(t)) =$$
$$\min_{\mathbf{n}_q(t)} \; c(\mathbf{n}_q(t)) + \rho\mathbb{E}\big[C_{q-1}\big(\mathbf{L}(t+1), \mathbf{x}(t+1), N(t+1)\big)\big],$$

which can be solved by the exhaustive search in the finite discrete action space [43]. According to [42], $\mathbf{n}_q(t)$ will finally converge to the actual $\mathbf{n}^*(t)$ as q increases.

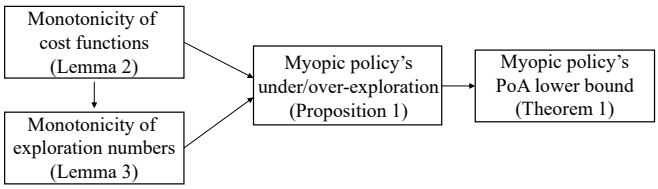

Fig. 2. Theoretical results of the myopic policy and the socially optimal policy in Section IV.

IV. Policies Comparison via PoA Analysis

In this section, we first analyze the monotonicity of the two policies with respect to hazard belief $x_i(t)$ and variance cost $V(\cdot)$. Then, we show that the myopic policy misses both exploration and exploitation over time, as compared to the social optimum, leading to PoA $\geq$ 2. This implies at least doubled total travel cost for all users and motivates our mechanism design in Section V.

We first analyze the monotonicity of cost functions under both policies in (12) and (13) below.

Lemma 2: The long-term cost functions (12) and (13) under both policies increase with expected travel latency $\mathbb{E}[\ell_i(t)|x_i'(t-1)]$, hazard state $x_i(t)$ and users' variance cost $V(\cdot)$ on stochastic path i.

The proof of Lemma 2 is given in Appendix B of the supplement material. As hazard belief $x_i(t)$ increases, the expected correlation coefficient $\mathbb{E}[\alpha_i(t)|x_i'(t)]$ will also increase, which incurs a larger travel cost for future users choosing stochastic path i. As variance cost $V(\cdot)$ enlarges, users at any time t face a greater error to estimate the travel latency on path i, which incurs a larger travel cost in (9).

Based on the monotonicity of cost functions in Lemma 2, we then analyze the monotonicity of their corresponding exploration numbers $n_i^{(m)}(t)$ and $n_i^*(t)$ in the next lemma.

Lemma 3: Both exploration numbers $n_i^{(m)}(t)$ and $n_i^*(t)$ decrease with $x_i(t)$ and $V(\cdot)$. While the difference $n_i^*(t) - n_i^{(m)}(t)$ increases with $x_i(t)$ but decreases with $V(\cdot)$.

The proof of Lemma 3 is given in Appendix C of the supplement material. Intuitively, as hazard belief $x_i(t)$ increases, the myopic policy becomes less willing to explore stochastic path i, which has a larger immediate travel cost than safe path 0, than the socially optimal policy. However, as $V(\cdot)$ enlarges, users' explorations will incur extra error costs, making both policies unwilling to explore. Lemma 3 also tells that longer expected latency $\mathbb{E}[\ell_i(t)|x_i'(t-1)]$, stronger hazard belief $x_i(t)$, and enlarged variance cost $V(\cdot)$ may hinder allocating more users to explore stochastic path i.

A. Exploration and Exploitation Comparison

Thanks to Lemma 3, now we are ready to analytically compare $n_i^{(m)}(t)$ with $n_i^*(t)$. Before that, we first define the myopic routing policy's under/over-exploration of stochastic path i as compared to the social optimum in the following.

Definition 2 (Under/over exploration): The myopic policy under-explores stochastic path i if $n_i^{(m)}(t) < n_i^*(t)$, and over-explores if $n_i^{(m)}(t) \geq n_i^*(t)$ at time t.

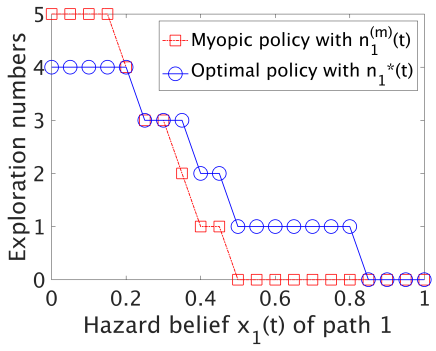

Fig. 3. Exploration numbers $n_1^*(t)$ under the socially optimal policy and $n_1^{(m)}(t)$ under the myopic policy versus hazard belief $x_1(t)$ in an illustrative two-path transportation network with $M = 1$.

Based on Definition 2, we next prove that the myopic policy misses both proper exploration and exploitation of stochastic path i, as hazard belief $x_i(t)$ dynamically changes over time.

Proposition 1: There exists a belief threshold $x_{th} \in (0, 1)$, such that the myopic policy will over-explore stochastic path i (with $n_i^{(m)}(t) \geq n_i^*(t)$) if $x_i(t) < x_{th}$, and will under-explore stochastic path i (with $n_i^{(m)}(t) < n_i^*(t)$) if $x_i(t) \geq x_{th}$, as compared to the socially optimal policy. This belief threshold x_{th} increases with $V(\cdot)$.

The proof of Proposition 1 is given in Appendix D of the supplement material. If there is a strong hazard belief $x_i(t) \geq x_{th}$ of $\alpha_i(t) = \alpha_H$, the myopic policy is not willing to explore stochastic path i due to longer latency. However, the socially optimal policy still allocates some users to learn possible α_L future others to exploit. In contrast, given weak $x_i(t) < x_{th}$, myopic users flock to stochastic path i without considering future congestion, while the social optimum may exploit safe path 0 to reduce congestion on path i. According to Lemma 3, as error cost $V(\cdot)$ increases, the socially optimal policy also becomes less willing to explore stochastic path i, leading to a belief threshold x_{th}.

To illustrate this, in Fig. 3, we simulate Fig. 1 using an illustrative basic two-path network with $M = 1$. We plot the myopic exploration number $n_1^{(m)}(t)$ in (10) and the optimal $n_1^*(t)$ derived by (13) versus hazard belief $x_1(t)$ of stochastic path 1. These two thresholds are very different as shown in Fig. 3. Here we set $\ell_0 = 15, \bar{N} = 5, \mathbb{E}[\ell_1(t-1)] = 20, \alpha_H = 1.5, \alpha_L = 0.05$ at current time t. Here $q_H(n_1(t))$ in (3) follows a normal distribution's CDF with mean 0.3 and variance 1, and we properly choose error function $V(n(t-1)) = \min\{\frac{10}{n(t-1)}, 20\}$. Given the belief threshold $x_{th} = 0.3$ here, if the hazard belief $x_1(t) < x_{th}$, we have the myopic exploration threshold $n_1^{(m)}(t)$ larger than the optimal $n_1^*(t)$ to over-explore stochastic path 1. If $x_1(t) \geq x_{th}$, the myopic exploration threshold satisfies $n_1^{(m)}(t)$ smaller than the optimal $n_1^*(t)$ to under-explore. This result is consistent with Proposition 1.

B. PoA Analysis

After comparing the two policies, we want to quantify the efficiency gap between their corresponding social costs. Define the price of anarchy (PoA) as the maximum ratio between the social cost in (12) under the myopic policy and the minimum cost in (13) under the socially optimal policy [40]:

$$\text{PoA}^{(m)} = \max_{\substack{\alpha_H, \alpha_L, \mathbf{x}(t), \ell_0, N(t) \\ \mathbf{L}(t), q_H, q_L, V(\cdot)}} \frac{C^{(m)}\big(\mathbf{L}(t), \mathbf{x}(t), N(t)\big)}{C^*\big(\mathbf{L}(t), \mathbf{x}(t), N(t)\big)}, \quad (14)$$

by searching all possible parameters and error function $V(\cdot)$ of the dynamic system. It is obvious that $\text{PoA}^{(m)} \geq 1$.

In the next theorem, we prove the lower bound of PoA.

Theorem 1: In the parallel traffic network with M stochastic paths, as compared to the minimum social cost in (13), the myopic policy in (12) results in:

$$\text{PoA}^{(m)} \geq \frac{2(1-\rho^{\Psi})}{2-\rho-\rho^{\Psi}}, \quad (15)$$

where $\Psi = 1 + \log_{\alpha_H}\left(M\frac{\left(\ell_0 - \frac{N_{min}}{M} - V(\frac{N_{min}}{M})\right)(\alpha_H - 1)}{\alpha_H N_{max}} + 1\right)$. The lower bound in (15) is larger than 2 as $\rho \to 1$, $V(\frac{N_{min}}{M}) \ll \ell_0$ and $\ell_0 \gg \alpha_H \frac{N_{max}}{M}$.

The proof of Theorem 1 is given in Appendix E of the supplement material. Inspired by Proposition 1, the worst case may happen when the myopic policy maximally under-explores stochastic paths with strong hazard belief $x_i(t)$. Initially, we set expected correlation coefficient $\mathbb{E}[\alpha_i(0)|x_i(0)] = 1$ and expected travel costs $c_i(0) = c_0(N_{max})$ for any stochastic path i. Then according to (10), myopic users always choose safe path 0 to make $n_i^{(m)}(t) = 0$. However, the socially optimal policy frequently asks $n_i^*(t) > 0$ of new user arrivals to explore stochastic path i to learn $\alpha_L = 0$, which greatly reduces future travel latency on this path. After that, all users will keep exploiting path i with $\alpha_i(t) = \alpha_L$, and the expected travel cost on this path may gradually increase to $c_i(\frac{N_{min}}{M}) = c_0(0)$ again after at least Ψ time slots. As $\Psi \to \infty$ (by setting small variance cost $V(\frac{N_{min}}{M}) \ll \ell_0$ and large latency $\ell_0 \gg \alpha_H \frac{N_{max}}{M}$) and $\rho \to 1$, we have $\text{PoA}^{(m)} \geq 2$.

We can also show that the $\text{PoA}^{(m)}$ lower bound in (15) decreases with variance cost $V(\cdot)$. This is consistent with Lemma 3 and Proposition 1 that $n_i^{(m)}(t)$ approaches to $n_i^*(t)$ as $V(\cdot)$ enlarges. Besides, if stochastic path number $M \to \infty$, Ψ approaches infinity, and the lower bound of PoA in (15) becomes 2, as more stochastic paths can negate the congestion cost caused by users' exploration. As users' myopic routing can at least double the social cost with $\text{PoA}^{(m)} \geq 2$, we are well motivated to design an efficient mechanism to regulate.

V. CHAR Mechanism with Learning Convergence

In this section, we aim to design an informational mechanism to regulate users' myopic routing. We first analyze $x_i(t)$'s learning convergence under both myopic and socially optimal policies. Based on this analysis, we further show that existing information-hiding and deterministic-recommendation mechanisms are far from efficient. Accordingly, we propose our combined hiding and probabilistic recommendation (CHAR) mechanism and prove its close-to-optimal efficiency.

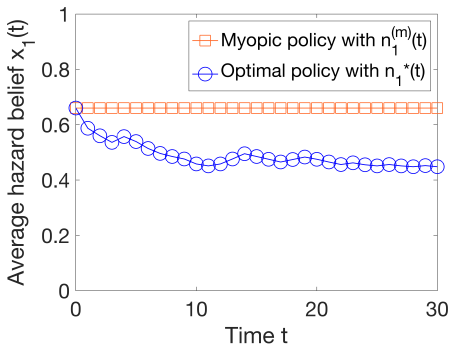

Fig. 4. Dynamics of average hazard belief $x_1(t)$ under both myopic and socially optimal policies from $t = 0$ to 30 in a two-path network with $M = 1$.

A. Learning Convergence Analysis

In the next lemma, we show that the myopic policy fails to ensure correct long-run convergence of hazard belief $x_i(t)$.

Lemma 4: Under the myopic policy with exploration number $n_i^{(m)}(t)$ on path i, the learned $x_i(t)$ is not guaranteed to converge to its real steady state $\bar{x}$ in (2) as $t \to \infty$.

The proof of Lemma 4 is given in Appendix F of the supplement material. Under the myopic policy, users only care about the immediate travel latency of each path. If the travel latencies of stochastic paths all satisfy $\mathbb{E}[\ell_i(t)|x_i(t)] > \ell_0$ with $\mathbb{E}[\alpha_i(t)|x_i(t)] \geq 1$, myopic user arrivals will never explore any stochastic path i and fail to update $x_i(t)$. In this case, hazard belief $x_i(t)$ may deviate a lot from its real steady state $\bar{x}$. The deviation of $x_i(t)$ also incurs extra costs to mislead future users. Next, we examine the learning convergence of $x_i(t)$ under the socially optimal policy.

Proposition 2: Under the socially optimal policy with exploration number $n_i^*(t)$, once $V(N_{min}) < \ell_0$, as $t \to \infty$, $x_i(t)$ is guaranteed to converge to its real steady state $\bar{x}$ in (2).

The proof of Proposition 2 is given in Appendix G of the supplement material. If $V(N_{min}) \geq \ell_0$, the huge variance cost discourages the system to explore and converge to the real expected belief $\bar{x}$ in any way. Note that even with $V(N_{min}) < \ell_0$, the myopic policy is not guaranteed to converge to real expected belief $\bar{x}$. While the socially optimal policy is willing to frequently explore stochastic path i to learn a low-hazard state α_L to reduce the expected travel latency for the followers. With long-term frequent exploration, hazard belief $x_i(t)$ can finally converge to its real steady state $\bar{x}$. The convergence to $\bar{x}$ also helps the system best decide routing and reduce social costs.

In Fig. 4, we run multiple experiment instances to show the average convergence of $x_1(t)$ under both policies in a finite time horizon $\{0, \cdots, 30\}$. Here we set $N(0) = \bar{N} = 7, \ell_0 = 100, x_1(0) = 0.66, \mathbb{E}[\ell_1(0)|x_1(0) = 0.6] = 110, n_1(0) = 0, \rho = 0.99, \alpha_H = 1.5, \alpha_L = 0.05$, and $\bar{\alpha} = 0.7$ with $\bar{x} = 0.45$ at initial time $t = 0$. Here $q_H(n_1(t))$ in (3) follows a normal distribution's CDF with mean 0.3 and variance 1, and we properly choose error function $V(n(t-1)) = \min\{\frac{10}{n(t-1)}, 20\}$. Under these parameter settings, the myopic policy never explores stochastic path 2, because the expected latency on path 2 keeps increasing under $\mathbb{E}[\alpha(t)|x(t-1)] = 1.007$. Hence, the myopic policy makes $x(t) = 0.66$ deviate from real but unknown $\bar{x} = 0.45$ in Fig. 4. However, the socially optimal policy still recommends some users to learn useful information on path 2, which ensures $x(t)$'s correct convergence to $\bar{x} = 0.45$. The simulation results above are consistent with Lemma 4 and Proposition 2.

B. Benchmark Informational Mechanisms Comparison

Before presenting our new mechanism, we first follow the mechanism design literature [44] to generally define the informational mechanism below.

Definition 3 (Informational mechanisms): An informational mechanism defines an incomplete information Bayesian game, where the social planner determines the disclosing level of variables based on the system state.

In practice, informational mechanisms are non-monetary and easier to implement as compared to pricing and side-payment mechanisms (e.g., [26]–[28]). Then we analyze the efficiency of two widely used informational mechanisms: information-hiding [29]–[31] and deterministic-recommendation [32]–[34]. Note that we also tried prior Bayesian persuasion mechanisms [18]–[20], which focus on one-shot games and cannot optimize long-run information learning.

Based on Proposition 2, if the platform hides all the information, i.e., $\mathbf{L}(t), \mathbf{x}(t)$ and $N(t)$, as [29]–[31], users can only estimate that each stochastic path has reached its steady state $\bar{x} \sim \mathbb{P}(\bar{x})$ in (2) and $\mathbb{E}[N(t)] = \bar{N}$ for any time t. As in [29], expected exploration number $n_i^{\emptyset} \in \{0, 1, \cdots, N(t)\}$ of stochastic path i under information-hiding becomes constant:

$$n_i^{\emptyset} = \min\left\{\frac{N(t)}{M}, \frac{\bar{N} + c_0(0) - \mathbb{E}_{\bar{x}\sim\mathbb{P}(\bar{x})}[c_i(0)|\bar{x}]}{M+1}\right\}, \quad (16)$$

where $c_0(\cdot)$ and $c_i(\cdot)$ are defined in (8) and (9), respectively. We next prove the infinite PoA caused by the hiding mechanism.

Lemma 5: If the platform hides all the information, i.e., $\mathbf{L}(t), \mathbf{x}(t)$ and $N(t)$, from users, the constant policy $n_i^{\emptyset}$ in (16) makes users either under- or over-explore stochastic path i as compared to the social optimum, leading to PoA $= \infty$.

The proof of Lemma 5 is given in Appendix H of the supplement material. If $\mathbb{E}_{\bar{x}\sim\mathbb{P}(\bar{x})}[c_i(\frac{N_{max}}{M})|\bar{x}] \leq c_0(0)$, all the $N(t)$ users without information will choose stochastic paths with less expected cost. However, the actual $\alpha_i(t)$ can be α_H and $\mathbb{E}[\ell_i(t)|x_i(t)] \gg \ell_0$, then users' maximum over-exploration makes PoA arbitrarily large. Similarly, if $\mathbb{E}_{\bar{x}\sim\mathbb{P}(\bar{x})}[c_i(0)|\bar{x}] > c_0(N_{max})$, all users will choose path 0, leading to the maximum under-exploration.

Next, we consider the existing deterministic recommendation mechanisms ([32]–[34]), which privately provide state-dependent deterministic recommendations to users while hiding other information, to prove the caused infinite PoA.

Lemma 6: The deterministic-recommendation mechanism makes PoA $= \infty$ in our system.

The proof of Lemma 6 is given in Appendix I of the supplement material. When a user receives recommendation $\pi(t) = i$, he only infers $n_i(t) \geq 1$ on path i. However, if

expected user number $\bar{N} \gg 1$, this information is insufficient to alter his posterior distribution of $\bar{x}$ from $\mathbb{P}(\bar{x})$. Consequently, the caused exploration number still approaches $n_i^{\emptyset}$ in (16), leading to PoA $= \infty$.

C. New CHAR Mechanism Design and Analysis

Inspired by Subsection V-B, the previous two mechanisms do not work in improving our system performance. Thus, we will properly combine them for a new mechanism design. To approach optimal policy $n_i^*(t)$ as much as possible, we dynamically select a number $N^{\emptyset}(t)$ of users to follow hiding policy $n_i^{\emptyset}$ in (16), while providing state-dependent probabilistic recommendations to the remaining $N(t) - N^{\emptyset}(t)$ users.

Now we are ready to propose our combined hiding and probabilistic recommendation (CHAR) mechanism, which contains two steps per time slot t. Under CHAR, let $n_i^{(\text{CHAR})}(t) \in \{0, \cdots, N(t)\}$ represent the exploration number of path i at t. Similar to (12), denote $C^{(\text{CHAR})}(\mathbf{L}(t), \mathbf{x}(t), N(t))$ to be the long-term social cost under $\mathbf{n}^{(\text{CHAR})}(t)$.

Definition 4 (CHAR mechanism): At any time t, in the first step, the platform randomly divides the $N(t)$ users into the hiding-group with $N^{\emptyset}(t)$ users and the recommendation-group with $N(t) - N^{\emptyset}(t)$ users, where the dynamic number $N^{\emptyset}(t)$ is the optimal solution to

$$\min_{N^{\emptyset}(t) \in \{0, \cdots, N(t)\}} C^{(\text{CHAR})}(\mathbf{L}(t), \mathbf{x}(t), N(t)), \tag{17}$$

$$\text{s.t.} \qquad n_i^{(\text{CHAR})}(t) = N^{\emptyset}(t) \cdot \frac{n_i^{\emptyset}}{N(t)} + (N(t) - N^{\emptyset}(t)) \mathbf{Pr}\big(\pi(t) = i | x_i(t)\big), \tag{18}$$

where

$$\mathbf{Pr}\big(\pi(t) = i | x_i(t)\big) = \begin{cases} p_L, & \text{if } x_i(t) < x_{th}, \\ p_H, & \text{if } x_i(t) \geq x_{th}, \end{cases} \tag{19}$$

with x_{th} is in Proposition 1 and any feasible p_L and p_H to satisfy $p_L \mathbb{P}(x_{th}) \geq p_H(1 - \mathbb{P}(x_{th}))$.

In the second step, the platform performs as follows:

- For the $N^{\emptyset}(t)$ hiding-group users, the platform hides all the past information from them.
- For the rest $N(t) - N^{\emptyset}(t)$ users in the recommendation-only group, the platform randomly recommends a path $\pi(t)$ to each user arrival by the following distribution:

$$\pi(t) = \begin{cases} i, & \text{with } \mathbf{Pr}(\pi(t) = i | x_i(t)) \text{ in (19)}, \\ 0, & \text{with } 1 - \sum_{i=1}^{M} \mathbf{Pr}(\pi(t) = i | x_i(t)). \end{cases} \tag{20}$$

According to Definition 4, the $N^{\emptyset}(t)$ users in the uniformly selected hiding-group rely on hiding policy $n_i^{\emptyset}$ in (16) to make routing decisions. For the recommendation-only group, given the always feasible p_L and p_H to satisfy $p_L \mathbb{P}(x_{th}) \geq p_H(1 - \mathbb{P}(x_{th}))$, if a user receives a recommendation $\pi(t) = i$, he infers that the posterior belief of low hazard $x_i(t) < x_{th}$ on path i is larger than the high harzard $x_i(t) \geq x_{th}$. Thus, this user will not deviate from the recommendation for a smaller expected travel cost. Similarly, if recommended $\pi(t) = 0$, the user can infer that any stochastic path i is more likely to have $x_i(t) \geq x_{th}$ and accordingly will choose safe path 0.

Given Bayesian incentive compatibility for all users, according to (18), there are respectively $N^{\emptyset}(t) \cdot n_i^{\emptyset}/N(t)$ users in the hiding-group and $(N(t) - N^{\emptyset}(t)) \cdot \mathbf{Pr}(\pi(t) = i | x_i(t))$ users in the recommendation-only group traveling on any path i. If stochastic path i has a good condition with $x_i(t) < x_{th}$ at time t, the recommendation probability in (19) is $\mathbf{Pr}\big(\pi(t) = i | x_i(t)\big) = p_L$. Then the exploration number of users on path i under our CHAR mechanism satisfies

$$n_i^{(\text{CHAR})} = N^{\emptyset}(t)\Big(\frac{n_i^{\emptyset}}{N(t)} - p_L\Big) + p_L N(t),$$

which depends on the choice of $N^{\emptyset}(t) \in \{0, \cdots, N(t)\}$. While if path i has a bad condition with $x_i(t) \geq x_{th}$, $n_i^{(\text{CHAR})}$ is similarly derived. To make $n_i^{(\text{CHAR})}(t)$ approach optimum $n_i^*(t)$ on each path i, the platform optimizes (17) to determine the optimal solution $N^{\emptyset}(t)$ in each time slot t.

Then we propose the following theorem to prove the close-to-optimal performance of our CHAR mechanism.

Theorem 2: Our CHAR mechanism in Definition 4 ensures Bayesian incentive compatibility for all users and regulates the dynamic routing system to achieve

$$\text{PoA}^{(\text{CHAR})} = 1 + \frac{1}{2(M+1)\big(1 + \frac{M}{\bar{N}} \cdot V\big(\frac{\bar{N}(2M+1)}{2M(M+1)}\big)\big)}, \tag{21}$$

which is always less than $\frac{5}{4}$ and cannot be further reduced by any informational mechanism.

The proof of Theorem 2 is given in Appendix J of the supplement material. On one hand, our CHAR mechanism recommends $n_i^{(\text{CHAR})} > n_i^{(m)}(t)$ users to explore stochastic path i of high belief $x_i(t) \geq x_{th}$. This successfully avoids myopic policy's zero-exploration in Theorem 1. On the other hand, our CHAR well controls the number of users on stochastic paths with bad conditions, avoiding benchmark mechanisms' maximum exploration in bad traffic conditions in Lemmas 5 and 6. Therefore, it efficiently reduces $\text{PoA}^{(m)} \geq 2$ caused by the myopic policy. Under our CHAR, the worst case occurs when users over-explore stochastic paths in good conditions ($\alpha_i(t) = \alpha_L$). We consider the scenario where the travel cost on any stochastic path i satisfies $c_i\big(\frac{N_{max}}{M}\big) < c_0(0)$ even if $\alpha_i(t) = \alpha_H$. Then all users will opt for stochastic paths, and no information incentives can curb their over-exploration. The achieved minimum PoA is derived in (21), which cannot be further reduced by any informational mechanism.

Our CHAR's PoA in (21) increases with the expected user number $\bar{N}$ and decreases with path number M. As $\bar{N} \to \infty$, $\text{PoA}^{(\text{CHAR})}$ converges to $1 + \frac{1}{2(M+1)}$ with minimum $\frac{5}{4}$. While if $M \to \infty$, $\text{PoA}^{(\text{CHAR})}$ approaches the optimum 1.

VI. Extensions of Parallel Transportation Network to Linear Path Graphs

In this section, we extend our congestion model and analysis from the parallel transportation network between two nodes O and D in Fig. 1 to more general linear path networks with multiple intermediate nodes and stochastic paths ([45]). In this generalized model, we first establish that the PoA lower

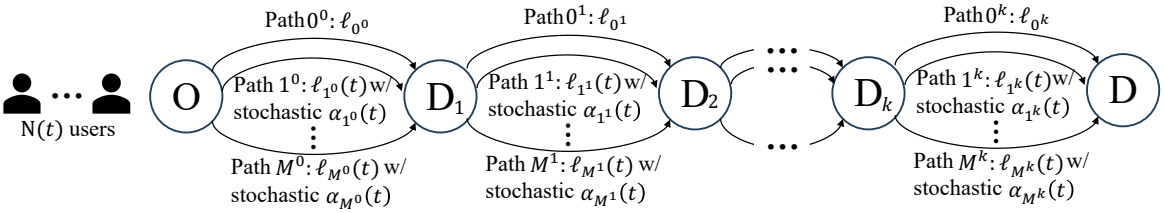

Fig. 5. Generalization from the parallel graph in Fig. 1 to a linear path graph: At the beginning of each time slot $t \in \{1, 2, \cdots\}$, $N_j(t)$ users arrive at any node $\mathrm{D}_j \in \{\mathrm{O}, \mathrm{D}_1, \cdots, \mathrm{D}_k\}$ selects one path from $M+1$ available paths to travel to the next node in this linear path graph. Among all the $M+1$ paths between intermediate nodes D_j and D_{j+1}, where $j \in \{0, \cdots, k\}$, path 0^j is safe while any path $i^j \in \{1^j, \cdots, M^j\}$ is stochastic.

bound caused by the myopic policy remains unchanged. Next, we prove that our CHAR mechanism continues to effectively reduce the PoA to its minimum possible value.

As depicted in Fig. 5, we consider a general linear path graph, which contains an order list of k intermediate nodes, denoted by $\mathrm{D}_1, \cdots, \mathrm{D}_k$, between origin O and destination D. For ease of exposition, we denote $\mathrm{D}_0 = \mathrm{O}$ and $\mathrm{D}_{k+1} = \mathrm{D}$. At each node D_j for $j \in \{0, 1, \cdots, k\}$, there exist one safe path 0^j and M stochastic paths $1^j, \cdots, M^j$ leading to the next node D_{j+1}. The safe path 0^j within the j-th subnetwork has a fixed travel latency ℓ_{0^j}. While for the other M^j stochastic paths, their correlation coefficients still alternate between a high-hazard state α_H and a low-hazard state α_L over time.

Within each subnetwork j from node D_j to D_{j+1} of this generalized path network, users traveling on stochastic path i^j will learn the traffic information there and update its hazard belief $x_{i^j}(t+1)$ for the platform as follows:

$$x_{i^j}(t+1) = x'_{i^j}(t) q_{HH}(t) + (1 - x'_{i^j}(t)) q_{LH}(t).$$

In this extended system model, a user arriving at node D_j not only needs to consider his current travel cost $c_{i^j}(n_{i^j}(t))$ in the j-th subnetwork, but also cares about his cost-to-go in the future subnetworks $\{j+1, \cdots, k\}$. This makes the myopic policy different from the one-shot decision in (10) for the parallel network in Fig. 1. Nonetheless, we still managed to obtain the PoA caused by the myopic routing policy in the following proposition.

Proposition 3: Under the general linear path graph in Fig. 5, as compared to the social optimum, the myopic policy achieves

$$\mathrm{PoA}^{(m)} \geq \frac{2(1-\rho^{\Psi})}{2-\rho-\rho^{\Psi}},$$

where $\Psi = 1 + \log_{\alpha_H}\left(M \frac{\left(\ell_{0^j} - \frac{N_{min}}{M} - V(\frac{N_{min}}{M})\right)(\alpha_H - 1)}{\alpha_H N_{max}} + 1\right)$. The lower bound in (15) is larger than 2 as $\rho \to 1$, $V(\frac{N_{min}}{M}) \ll \ell_{0^j}$ and $\ell_{0^j} \gg \alpha_H \frac{N_{max}}{M}$ for any $j \in \{0, \cdots, k\}$.

The proof of Proposition 3 is given in Appendix K of the supplement material. In the linear path network illustrated in Fig. 5, we set an expected correlation coefficient $\mathbb{E}[\alpha_{i^j}(0) | x_{i^j}(0)] = 1$ and expected travel costs $c_{i^j}(0) = c_{0^j}(N_{max})$ for any stochastic path i^j. In the worst-case scenario with $\rho = 1$, the myopic policy still leads all users to choose the safe path 0^j at any node D_j. While the socially optimal policy always recommends $n^*_{i^j}(t) > 0$ users to explore stochastic path i^j to potentially learn possible $\alpha_L = 0$ there. After that, all users keep exploiting stochastic path i^j with $\alpha_{i^j}(t) = \alpha_L$, resulting in PoA > 2. It is noteworthy that in this case, both the myopic policy and the socially optimal policy just repeat their decision-making processes, as in (10) and (13), for $k+1$ times from origin O to destination D. Consequently, the resulting PoA remains the same as that of the parallel path network in Theorem 1.

Motivated by Proposition 3, we next examine our CHAR mechanism's performance for this generalized system. In the following, we consider $\rho = 1$ in the worst-case scenario to regulate myopic routing. Let vectors $\mathbf{L}^j(t)$ and $\mathbf{x}^j(t)$ summarize the latencies and beliefs of all the M stochastic paths within the j-th subnetwork. Denote by $N^j(t)$ the arriving user number at node D^j at time t. We first propose the following lemma to check users' decision-making without information sharing.

Lemma 7: If the platform hides all the information of the linear path graph in Fig. 5 from all users at any node D_j, i.e., $\mathbf{L}^j(t), \mathbf{x}^j(t)$ and $N^j(t)$ at any time t, the expected exploration number of stochastic path i^j within the j-th subnetwork equals

$$n^{\emptyset}_{i^j} = \min\left\{\frac{N^j(t)}{M}, \frac{\bar{N} + c_{0^j}(0) - \mathbb{E}_{\bar{x}\sim\mathbb{P}(\bar{x})}[c_{i^j}(0)|\bar{x}]}{M+1}\right\}. \tag{22}$$

The proof of Lemma 7 is given in Appendix L of the supplement material. Intuitively, without any information, users can only estimate that each stochastic path i^j has reached its steady state $\bar{x}$ in (2) and $\mathbb{E}[N^j(t)] = \bar{N}$ for any time t. Under the steady state, for any user at node D_j, his expected cost-to-go for the future subnetworks $\{j+1, \cdots, k\}$ is unaffected by his current path choice. Therefore, he only needs to focus on the current subnetwork j, as (16) in the parallel network, to make his routing decision. This leads to the total exploration number $n^{\emptyset}_{i^j}$ in (22) for path i^j.

Thanks to Lemma 7, we only need to focus on a single subnetwork $\{\mathrm{D}_j, \mathrm{D}_{j+1}\}$ to regulate each user arrival at node D_j. As a result, our CHAR mechanism for the parallel network in Definition 4 still works. Subsequently, we prove its PoA result in the following theorem.

Theorem 3: Under the linear path graph in Fig. 5, our CHAR mechanism in Definition 4 regulates the dynamic system to achieve the same PoA as (21), which is always less than $\frac{5}{4}$.

The proof of Theorem 3 is given in Appendix M of the supplement material. Since our CHAR mechanism only needs to regulate users within each subnetwork j, the achieved PoA within each subnetwork is equal to (21), as in the

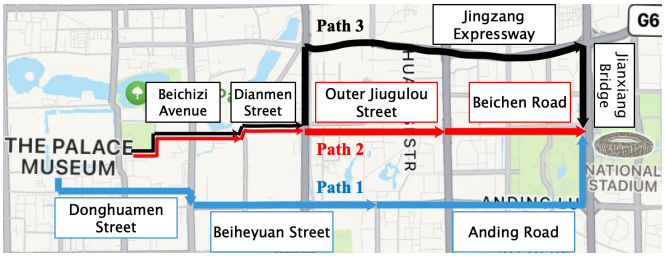

Fig. 6. A popular hybrid road network consisting of three path choices from the Palace Museum to the National Stadium in Beijing, China. In this hybrid network, each path has a series of roads, where Path 2 and Path 3 overlap on Beichizzi Avenue and Dianmen Street, making it different from the linear path graph in Fig. 5. On public holidays, many visitors randomly leave the Palace Museum and choose a path to travel to the National Stadium.

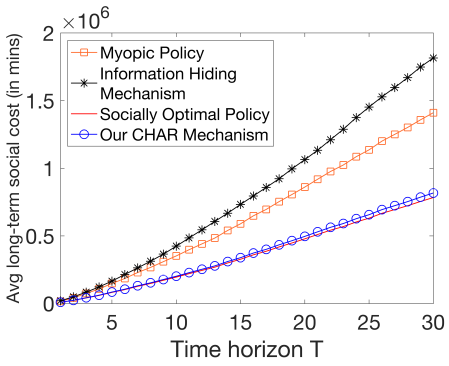

Fig. 7. Average long-term costs (in minutes) under the myopic policy, information hiding mechanism ([29]–[31]), the socially optimal policy, and our CHAR mechanism versus time horizon $T \in \{0, \cdots, 30\}$.

parallel transportation network. When considering all $k+1$ subnetworks, the PoA remains unchanged.

VII. Experiment Validation Using Real Datasets

In addition to the worst-case PoA analysis, in this section, we further conduct experiments to evaluate our CHAR's average performance versus the myopic policy used by Waze and Google Maps and the popular information hiding mechanism used in existing congestion game literature ([29]–[31]). To further practicalize our dynamic congestion model in (1), we sample peak hours' real-time traffic congestion data in Beijing, China on public holidays using BaiduMap ([46]), and further extend our parallel transportation network in Fig. 1 and linear path graph in Fig. 5 to a hybrid road network in Fig. 6.

As depicted in Fig. 6, we consider a popular hybrid road network from the Palace Museum to the National Stadium on public holidays, as many visitors randomly leave the Palace Museum and travel to the National Stadium. This hybrid network consists of three path choices, each with a series of safe and stochastic roads:

- Path 1: Donghuamen Street, Beiheyuan Street, and Anding Road.
- Path 2: Beichizi Avenue, Dianmen Street, Outer Jiugulou Street and Beichen Road.
- Path 3: Beichizi Avenue, Dianmen Street, Jingzang Expressway, and Jianxiang Bridge.

Here Path 2 and Path 3 overlap on Beichizzi Avenue and Dianmen Street, making it different from the linear path graph in Fig. 5.

To train a practical congestion model with travel latency, we mine and analyze many data about the real-time traffic status values of the nine roads from [46], which can dynamically change every 5 minutes in the peak-hour periods. We validate from the dataset that the traffic conditions of Donghuamen Street and Beiheyuan Street on Path 1, Beichizi Avenue on both Path 2 and Path 3, and Jianxiang Bridge on Path 3 can be well approximated as Markov chains with two discretized states (high and low traffic states) as in (2), while the other five roads tend to have deterministic/safe conditions. Similar to [47], [48], we employ the hidden Markov model (HMM) approach to train the transition probability matrices for the four stochastic road segments using 432 statuses with 5 minutes per time slot, which suffice to achieve high accuracy. Based on the transition probability matrices, we further calculate the steady-state belief $\bar{x}$'s of the four stochastic roads in Fig. 6:

$$\bar{x}_{\text{Donghuamen}} = 0.388,\ \bar{x}_{\text{Beihe}} = 0.106,$$
$$\bar{x}_{\text{Beichizi}} = 0.192,\ \bar{x}_{\text{Jianxiang}} = 0.936.$$

Then these steady-state beliefs can be used in the congestion model for our experiments.

Besides the above congestion model, based on the travel latency data in [46], we calculate the long-term average correlation coefficients $\alpha_H \approx 1.3$ and $\alpha_L \approx 0.3$, by assuming a linear latency function $f(\cdot)$ in (1). Then we use the traffic flow data from Baidu Map to calculate the mean arrival car number is $N = 121$ every 5 minutes at the gateway of the Palace Museum, with a standard deviation of 12.33. Given the mean arrival car number, we set $V(n(t-1)) = \min\{\frac{100}{n(t-1)}, 200\}$.

Based on the derived system parameters above, we are now ready to conduct our experiments. In the myopic policy, each user estimates the travel cost of any path choice, by aggregating the costs of its road segments, to select the one that minimizes his own travel cost, whereas the socially optimal policy seeks the path choice that minimizes the collective long-term social cost for all users. Our CHAR mechanism optimizes the user number of the hiding-group $N^{\emptyset}(t)$ in (17) and recommends path $\pi(t)$ to each user of the recommendation-only group in (20). This is achieved by combining the actual hazard beliefs of all stochastic road segments on each path choice. As users' costs are proportional to their travel delays, we conduct 50 experiments, each with 31 time slots (equivalent to 155 minutes), to calculate their average long-term social costs (in minutes). Here the initial hazard beliefs are approximated to $0.5, 0.2, 0.3$, and 0.8 for Donghuamen Street, Beiheyuan Street, Beichizi Avenue, and Jianxiang Bridge, respectively.

In the first experiment, we fix discount factor $\rho = 0.98$ to compare the average long-term social costs under the myopic policy, information hiding mechanism, the socially optimal policy, and our CHAR mechanism versus time horizon $T \in \{0, \cdots, 30\}$. Fig. 7 depicts that our CHAR mechanism has less than 5% efficiency loss from the social optimum for any time horizon T, while the myopic policy and the

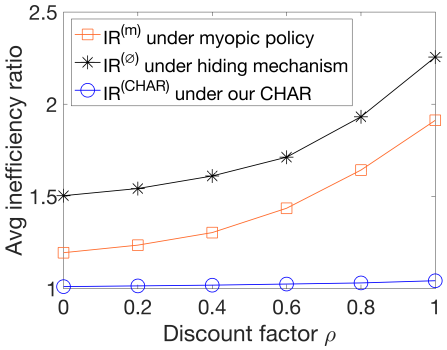

Fig. 8. Average long-term inefficiency ratios $\text{IR}^{(m)}$, $\text{IR}^{(\emptyset)}$, and $\text{IR}^{(\text{CHAR})}$ caused by the myopic policy, information-hiding mechanism, and our CHAR mechanism (as compared to the socially optimal policy) versus discount factor $\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

information hiding mechanism cause around 90% and 125% efficiency losses, respectively.

Next, we define

$$\text{IR}^{(m)} = \frac{\mathbb{E}[C^{(m)}(\mathbf{L}(t), \mathbf{x}(t), N(t))]}{\mathbb{E}[C^{*}(\mathbf{L}(t), \mathbf{x}(t), N(t))]}$$

to be the long-term average inefficiency ratio of the total travel cost $C^{(m)}(\mathbf{L}(t), \mathbf{x}(t), N(t))$ under the myopic policy in (12), as compared to $C^{*}(\mathbf{L}(t), \mathbf{x}(t), N(t))$ under the socially optimal policy in (13). Similarly, define

$$\text{IR}^{(\emptyset)} = \frac{\mathbb{E}[C^{(\emptyset)}(\mathbf{L}(t), \mathbf{x}(t), N(t))]}{\mathbb{E}[C^{*}(\mathbf{L}(t), \mathbf{x}(t), N(t))]},$$

$$\text{IR}^{(\text{CHAR})} = \frac{\mathbb{E}[C^{(\text{CHAR})}(\mathbf{L}(t), \mathbf{x}(t), N(t))]}{\mathbb{E}[C^{*}(\mathbf{L}(t), \mathbf{x}(t), N(t))]}$$

to be the average inefficiency ratios caused by the information hiding mechanism and our CHAR mechanism, respectively. Then in the second experiment, we fix time horizon $T = 30$ to compare the average inefficiency ratios $\text{IR}^{(m)}, \text{IR}^{(\emptyset)}$ and $\text{IR}^{(\text{CHAR})}$ caused by the myopic policy, information hiding mechanism, and our CHAR mechanism versus discount factor $\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

Fig. 8 tells that all the inefficiency ratios increase with discount factor ρ, which is aligned with Theorem 1. In the worst case with $\rho = 1$, our CHAR mechanism obviously reduces $\text{IR}^{(m)} > 1.9$ and $\text{IR}^{\emptyset} > 2.2$ under the myopic policy and hiding mechanism to the close-to-optimal ratio $\text{IR}^{(\text{CHAR})} < 1.1$, which is consistent with Theorem 2. Note that even in the best case with $\rho = 0$, both myopic policy and information hiding mechanism perform poor with $\text{IR}^{(m)} > 1.2$ and $\text{IR}^{\emptyset} > 1.5$, while our CHAR mechanism makes $\text{IR}^{(\text{CHAR})} < 1.05$.

VIII. Conclusion

In this paper, we focus on dynamic congestion games with endogenous time-varying conditions to analyze and regulate human-in-the-loop learning. In a typical parallel routing network, our analysis shows that the myopic routing policy (used by Google Maps and Waze) leads to severe under-exploration of stochastic paths. This results in a price of anarchy (PoA) greater than 2, as compared to the social optimum achieved through optimal exploration-exploitation tradeoff. Besides, the myopic policy fails to ensure the correct learning convergence about users' traffic hazard beliefs. To address this, we first show that existing information-hiding mechanisms and deterministic path-recommendation mechanisms in Bayesian persuasion literature do not work with even PoA $= \infty$. Accordingly, we propose a new combined hiding and probabilistic recommendation (CHAR) mechanism to hide all information from a selected user group and provide state-dependent probabilistic recommendations to the other user group. Our CHAR successfully ensures PoA less than $\frac{5}{4}$, which cannot be further reduced by any other informational mechanism. Besides parallel networks, we further extend our analysis and CHAR to more general linear path graphs with multiple intermediate nodes, and we prove that the PoA results remain unchanged. Additionally, we carry out experiments with real-world datasets to validate the close-to-optimal performance of our CHAR mechanism.

In the future, we consider generalizing our human-in-the-loop learning beyond transportation applications, by considering stochastic queueing service systems (e.g., hospitals and restaurants). In queueing systems, many selfish users (e.g., customers for dining) are unwilling to explore new restaurants to learn useful information (e.g., tastes and queue length) for future customers. Thus, we need to design an efficient mechanism to regulate their exploration and exploitation of different servers.

References

[1] H. Li and L. Duan, "Distributed Learning for Dynamic Congestion Games," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2024.

[2] H. Zhang, H. Ge, J. Yang, and Y. Tong, "Review of vehicle routing problems: Models, classification and solving algorithms," *Archives of Computational Methods in Engineering*, vol. 29, no. 1, pp. 195–221, 2022.

[3] T. Haselton, "How to use Google Waze," https://www.cnbc.com/2018/11/13/how-to-use-google-waze-for-directions-and-avoiding-traffic.html, 2018.

[4] FHWA, "Crowdsourcing for advancing operations," https://www.fhwa.dot.gov/innovation/everydaycounts/edc_6/crowdsourcing.cfm, 2024.

[5] Y. Zhu and K. Savla, "Information design in non-atomic routing games with partial participation: Computation and properties," *IEEE Transactions on Control of Network Systems*, 2022.

[6] S. Vasserman, M. Feldman, and A. Hassidim, "Implementing the wisdom of waze," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[7] H. Liu, K. Liu, and Q. Zhao, "Learning in a changing world: Restless multiarmed bandit with unknown dynamics," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1902–1916, 2012.

[8] A. Slivkins *et al.*, "Introduction to multi-armed bandits," *Foundations and Trends® in Machine Learning*, vol. 12, no. 1-2, pp. 1–286, 2019.

[9] S. Gupta, S. Chaudhari, G. Joshi, and O. Yağan, "Multi-armed bandits with correlated arms," *IEEE Transactions on Information Theory*, vol. 67, no. 10, pp. 6711–6732, 2021.

[10] A. Bozorgchenani, S. Maghsudi, D. Tarchi, and E. Hossain, "Computation offloading in heterogeneous vehicular edge networks: On-line and off-policy bandit solutions," *IEEE Transactions on Mobile Computing*, vol. 21, no. 12, pp. 4233-4248, 2021.

[11] H. Wang, Y. Yang, E. Wang, W. Liu, Y. Xu, and J. Wu, "Truthful user recruitment for cooperative crowdsensing task: A combinatorial multi-armed bandit approach," *IEEE Transactions on Mobile Computing*, vol. 22, no. 7, pp. 4314-4331, 2022.

[12] F. Li, D. Yu, H. Yang, J. Yu, K. Holger, and X. Cheng, "Multi-armed-bandit-based spectrum scheduling algorithms in wireless networks: A survey," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 24-30, 2020.

[13] C. Shi and C. Shen, "Federated multi-armed bandits," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 11, 2021, pp. 9603–9611.
[14] L. Yang, Y.-Z. J. Chen, S. Pasteris, M. Hajiesmaili, J. Lui, and D. Towsley, "Cooperative stochastic bandits with asynchronous agents and constrained feedback," *Advances in Neural Information Processing Systems*, vol. 34, pp. 8885–8897, 2021.
[15] L. Yang, Y.-Z. J. Chen, M. H. Hajiemaili, J. C. Lui, and D. Towsley, "Distributed bandits with heterogeneous agents," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 200–209.
[16] J. Zhu and J. Liu, "Distributed multi-armed bandits," *IEEE Transactions on Automatic Control*, 2023.
[17] C. Meng and A. Markopoulou, "On routing-optimal networks for multiple unicasts," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 111–115.
[18] S. Das, E. Kamenica, and R. Mirka, "Reducing congestion through information design," in *2017 55th annual allerton conference on communication, control, and computing (allerton)*. IEEE, 2017, pp. 1279–1284.
[19] E. Kamenica, "Bayesian persuasion and information design," *Annual Review of Economics*, vol. 11, pp. 249–272, 2019.
[20] Y. Mansour, A. Slivkins, V. Syrgkanis, and Z. S. Wu, "Bayesian exploration: Incentivizing exploration in bayesian games," *Operations Research*, vol. 70, no. 2, pp. 1105–1127, 2022.
[21] Y. Babichenko, I. Talgam-Cohen, H. Xu, and K. Zabarnyi, "Regret-minimizing bayesian persuasion," *Games and Economic Behavior*, vol. 136, pp. 226–248, 2022.
[22] S. Gollapudi, K. Kollias, C. Maheshwari, and M. Wu, "Online learning for traffic navigation in congested networks," in *International Conference on Algorithmic Learning Theory*. PMLR, 2023, pp. 642–662.
[23] Y. Yang and G. I. Webb, "Discretization for naive-bayes learning: managing discretization bias and variance," *Machine learning*, vol. 74, no. 1, pp. 39–74, 2009.
[24] F. Meunier and N. Wagner, "Equilibrium results for dynamic congestion games," *Transportation Science*, vol. 44, no. 4, pp. 524–536, 2010.
[25] G. Carmona and K. Podczeck, "Pure strategy nash equilibria of large finite-player games and their relationship to non-atomic games," *Journal of Economic Theory*, vol. 187, p. 105015, 2020.
[26] Z. Zheng, S. Yang, J. Xie, F. Wu, X. Gao, and G. Chen, "On designing strategy-proof budget feasible online mechanisms for mobile crowdsensing with time-discounting values," *IEEE Transactions on Mobile Computing*, vol. 21, no. 6, pp. 2088–2102, 2020.
[27] B. L. Ferguson, P. N. Brown, and J. R. Marden, "The effectiveness of subsidies and tolls in congestion games," *IEEE Transactions on Automatic Control*, vol. 67, no. 6, pp. 2729–2742, 2022.
[28] F. Li, Y. Chai, H. Yang, P. Hu, and L. Duan, "The effectiveness of subsidies and tolls in congestion games," *IEEE/ACM Transactions on Networking*, 2024.
[29] H. Tavafoghi and D. Teneketzis, "Informational incentives for congestion games," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2017, pp. 1285–1292.
[30] J. Wang and M. Hu, "Efficient inaccuracy: User-generated information sharing in a queue," *Management Science*, vol. 66, no. 10, pp. 4648–4666, 2020.
[31] F. Farhadi and D. Teneketzis, "Dynamic information design: a simple problem on optimal sequential information disclosure," *Dynamic Games and Applications*, vol. 12, no. 2, pp. 443–484, 2022.
[32] Y. Li, C. Courcoubetis, and L. Duan, "Recommending paths: Follow or not follow?" in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 928–936.
[33] M. Wu and S. Amin, "Learning an unknown network state in routing games," *IFAC-PapersOnLine*, vol. 52, no. 20, pp. 345–350, 2019.
[34] H. Li and L. Duan, "When congestion games meet mobile crowdsourcing: Selective information disclosure," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 5, 2023, pp. 5739–5746.
[35] I. Kremer, Y. Mansour, and M. Perry, "Implementing the "wisdom of the crowd"," *Journal of Political Economy*, vol. 122, no. 5, pp. 988–1012, 2014.
[36] X. Ban, R. Herring, P. Hao, and A. M. Bayen, "Delay pattern estimation for signalized intersections using sampled travel times," *Transportation Research Record*, vol. 2130, no. 1, pp. 109–119, 2009.
[37] I. Alam, D. M. Farid, and R. J. Rossetti, "The prediction of traffic flow with regression analysis," in *Emerging Technologies in Data Mining and Information Security*. Springer, 2019, pp. 661–671.
[38] M. Venanzi, A. Rogers, and N. R. Jennings, "Crowdsourcing spatial phenomena using trust-based heteroskedastic gaussian processes," in *First AAAI Conference on Human Computation and Crowdsourcing*, 2013.
[39] A. L. Smith and S. S. Villar, "Bayesian adaptive bandit-based designs using the gittins index for multi-armed trials with normally distributed endpoints," *Journal of Applied Statistics*, vol. 45, no. 6, pp. 1052–1076, 2018.
[40] T. Roughgarden, *Selfish routing and the price of anarchy*. MIT press, 2005.
[41] D. Bertsimas and J. Tsitsiklis, "Introduction to linear optimization," Athena Scientific Belmont, MA, 1997
[42] S. Ross, "Introduction to stochastic dynamic programming," *Academic press*, 2014.
[43] F. Yao, "Efficient dynamic programming using quadrangle inequalities," *Proceedings of the twelfth annual ACM symposium on Theory of computing*, pp. 429–435, 1980.
[44] D. Fudenberg and J. Tirole, *Game theory*. MIT press, 1991.
[45] H. J. Broersma and C. Hoede, "Path graphs," *Journal of graph theory*, vol. 13, no. 4, pp. 427–444, 1989.
[46] B. BaiduMap, "Baidu maps open platform," https://lbsyun.baidu.com/faq/api?title=webapi/traffic-roadseek, 2023.
[47] S. R. Eddy, "Profile hidden markov models." *Bioinformatics (Oxford, England)*, vol. 14, no. 9, pp. 755–763, 1998.
[48] Z. Chen, J. Wen, and Y. Geng, "Predicting future traffic using hidden markov models," in *2016 IEEE 24th international conference on network protocols (ICNP)*. IEEE, 2016, pp. 1–6.

Hongbo Li (S'24) received the B.S. degree in Electronics and Electric Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2019. He is working toward the Ph.D. degree with the Pillar of Engineering Systems and Design, Singapore University of Technology and Design (SUTD). Currently, he is a Visiting Scholar at The Ohio State University. His research interests include game theory and mechanism design, networked AI, and machine learning theory.

Lingjie Duan (S'09-M'12-SM'17) received the Ph.D. degree from The Chinese University of Hong Kong in 2012. He is an Associate Professor at the Singapore University of Technology and Design (SUTD) and is an Associate Head of Pillar (AHOP) of Engineering Systems and Design. In 2011, he was a Visiting Scholar at University of California at Berkeley, Berkeley, CA, USA. His research interests include network economics and game theory, network security and privacy, energy harvesting wireless communications, and mobile crowdsourcing. He is an Associate Editor of IEEE/ACM Transactions on Networking and IEEE Transactions on Mobile Computing. He was an Editor of IEEE Transactions on Wireless Communications and IEEE Communications Surveys and Tutorials. He also served as a Guest Editor of the IEEE Journal on Selected Areas in Communications Special Issue on Human-in-the-Loop Mobile Networks, as well as IEEE Wireless Communications Magazine. He is a General Chair of WiOpt 2023 Conference and is a regular TPC member of some other top conferences (e.g., INFOCOM, MobiHoc, SECON). He received the SUTD Excellence in Research Award in 2016 and the 10th IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2015.