

A large-scale multi-rider matching problem with renegeing passengers: single source case

Jing Fu ^a Lele Zhang ^b Kai Huang ^c and Zhiyuan Liu ^d

^a*School of Engineering, RMIT University, VIC 3000, Australia,*

^b*School of Mathematics and Statistics, The University of Melbourne, VIC 3010, Australia,*

^c*Department of Civil Engineering, Monash University, VIC 3800, Australia,*

^d*School of Transportation, Southeast University, Nanjing, China*

Email: jing.fu@rmit.edu.au

Abstract: This paper studies a large-scale ride-matching problem with a single transport hub with a large number of passengers who are different in terms of destinations and travel preferences. Passengers with similar itineraries can match each other and share the same vehicle for their potentially different destinations; and renegeing passengers, who become impatient and leave the service system after waiting long time for shared rides, are considered in our model. We aim to maximize the long-run average revenue of the ride service vendor, which is defined as the difference between the long-run average reward earned by providing ride services and the long-run average penalty incurred by renegeing passengers. The problem is complicated by its scale, the heterogeneity of passengers, and the renegeing behaviors. To this end, we formate the ride-matching problem as a specific Markov decision process and propose a scalable ride-matching policy, referred to as Bivariate Index (BI) policy. Through extensive numerical simulations for systems with real-world travel demands, it is demonstrated that BI significantly outperforms baseline policies.

Keywords: *Restless bandits, ride-matching, renegeing passengers*

1 INTRODUCTION

The rider-matching modeling and optimization have been widely studied for years. Most of existing work focused on services with *single-rider* shared vehicles, for which a vehicle is shared by only one passenger and one driver. It followed studies such as Lee, Wang, Cheu & Teo (2004), Lee & Savelsbergh (2015), Levin *et al.* (2017). Detailed surveys for single-rider sharing models can be found in Agatz *et al.* (2012), Wang & Yang (2019). Studies also explored *multi-rider* matching problems, where each vehicle is potentially shared by different (groups of) passengers, have become popular in recent years Hong *et al.* (2017), Dong *et al.* (2018), Lokhandwala & Cai (2018).

We focus on a specific variant of multi-rider matching: dispatching passengers in a large transport hub, such as an airport and a train station, in a real-time manner. All passengers share the same source of travel routes (that is, the transport hub) but potentially different destinations and ride preferences. The problems with one travel source have been widely studied in literature (see Clarke & Freilich (1987), Lee, Wu & Wang (2004), Lee *et al.* (2005), Liu *et al.* (2018), Zhong & Liu (2020)) and are appropriate for systems with a large, busy transport terminal that generates high volume of ride demands and is located far away from city centers. The ride-sharing service vendors will *match* the waiting passengers with similar destinations and preferences, and assign vehicles to send them away. The vehicle drivers are contracted with the service vendors and provide taxi-like transport to multiple groups of passengers with similar itineraries. Lyu *et al.* (2019) indicated that, in this case, the shared rides can reduce detour distances and hence reduce the travel costs for these drivers. From the view of the service vendor, such a matching problem is complicated by the heterogeneity of arrived passengers and the extremely large population seeking for ride-sharing services at the transport hub.

Our problem is significantly complicated by the heterogeneity of passengers, their renegeing behaviors and the large scale of systems. A key step of this paper is to model the ride-matching problem as a Markov decision process (MDP) and approximate its optimality by applying the Whittle relaxation technique (see Whittle (1988)), which was originally proposed for the restless multi-armed bandit problem (RMABP). It has been proved to be PSPACE-hard in general (see Papadimitriou & Tsitsiklis (1999)), and people resort to scalable heuristics with guaranteed performance degradation. In particular, the problem is intractable as the state space of its underlying MDP increases exponentially in the number of distinct passengers.

Following the idea of Whittle relaxation (see Whittle (1988)), we are able to propose a scalable policy which approximates optimality of the ride-matching problem. We propose a scalable policy for the ride-matching problem in a large transport hub with hundreds of thousands of daily passengers, who have potentially different destinations and travel preferences. The effectiveness of the proposed policy is demonstrated through extensive numerical simulations with real-world settings: compared to the baseline policies, it achieves more than 20% higher average rewards. To the best of our knowledge, no existing studies for the multi-rider matching problem with enabled renegeing behaviors can be directly applied here.

2 A RIDE-MATCHING PROBLEM

We use \mathbb{N}_+ and \mathbb{N}_0 as the sets of positive and non-negative integers, and for any $N \in \mathbb{N}_+$, let $[N]$, represent the set $\{1, 2, \dots, N\}$ with $[0] = \emptyset$ and $[1] = \{1\}$. Let \mathbb{R}, \mathbb{R}_+ and \mathbb{R}_0 be the set of all, positive and non-negative reals, respectively.

2.1 System Model

Consider a ride-sharing system with L different *types* of customers awaiting transport service in a traveling center/airport/train station. Customers of the same type will destine to the same geographical region, such as a neighborhood or a suburb, in which all addresses are reachable from each other. Define the traveling destination for customer type $\ell \in [L]$ as a random point taking values representing locations in the associated geographical region, and refer to it as the *random destination* of type ℓ . Customers of different types differ from others in terms of random destinations, arrival rates, *renegeing rates* and potentially other travel preferences. The renegeing rate describes the probability that customers become impatient and depart a ride service after joining it without getting served. Customers with higher renegeing rates will become impatient more easily and are more sensitive to waiting time.

We refer to a customer of type $\ell \in [L]$ as an ℓ -customer. Here, a customer is a logic term that represents either one person or a group/family of people. Consider J *ride-sharing services* (RS services), each of which is specified by the travel preferences, including routes, of two types (not necessarily distinct) of customers that are able to share the same vehicle. Let $\mathcal{L}(j) \in 2^{[L]}$ represent the set of such two customer types associated with

RS service $j \in [J]$. RS service $j \in [J]$ offers a vehicle that drives an ℓ_1 and an ℓ_2 -customer, $\ell_1, \ell_2 \in \mathcal{L}(j)$, to their destinations through routes $r_1 \in \mathcal{R}_{\ell_1}$ and $r_2 \in \mathcal{R}_{\ell_2}$. The r_1 and r_2 are such random trajectories that, with a positive probability, one is covered by the other: the two customers can go through with one vehicle along some instances of the random routes. Note that, for such $\ell_1, \ell_2 \in \mathcal{L}(j)$, we allow the case with $|\mathcal{L}(j)| = 1$ and $\ell_1 = \ell_2$: two customers of the same type will share one vehicle. Let \mathcal{F}_ℓ represent the set of RS services able to serve an ℓ -customer.

The system controller (the vendor) will assign an ℓ -customer, once they arrived, to be *served* by one of such services $j \in \mathcal{F}_\ell$. For a RS service $j \in [J]$ and $\ell \in \mathcal{L}(j)$, let $N_{\ell,j}(t) \in \mathbb{N}_0$ represent the number of ℓ -customers that are waiting to be served by RS service j at time $t \geq 0$, that is, the number of ℓ -customers waiting in RS service j . When an ℓ -customer arrives at time $t \geq 0$, the system controller decides to assign it to a RS service $j \in \mathcal{F}_\ell$, and this assignment is not changed until the customer be served or becomes impatient. When an ℓ -customer arrives at time $t \geq 0$ and the system controller decides to assign it to a RS service $j \in \mathcal{F}_\ell$:

- if there is an ℓ -customer waiting there (that is, $N_{\ell,j}(t) > 0$) with $|\mathcal{L}_j| = 1$ (that is, the RS service is for customers of the same type), then both customers of type ℓ will leave immediately with a shared vehicle and $N_{\ell,j}(t)$ decrements by one;
- if there is an ℓ' -customer waiting there (that is, $N_{\ell',j}(t) > 0$) with $\ell' \in \mathcal{L}(j)$ and $\ell' \neq \ell$, then both customers will leave immediately with a shared vehicle and $N_{\ell',j}(t)$ decrements by one;
- otherwise, the number of waiting ℓ -customers in RS service j , $N_{\ell,j}(t)$, increments by one.

In this context, for any $t \geq 0$, $j \in [J]$ and $\ell, \ell' \in \mathcal{L}(j)$, if $\ell \neq \ell'$, then either $N_{\ell,j}(t) = 0$ or $N_{\ell',j}(t) = 0$; if $|\mathcal{L}(j)| = 1$ (so that $\ell = \ell'$), $N_{\ell,j}(t)$ is either 0 or 1.

Also, an ℓ -customer waiting for RS service $j \in \mathcal{F}_\ell$ may become *impatient* due to a long waiting time and prefer an immediate leave with a non-shared vehicle or another transport mode, such as a bus, though it incurs higher charge or inconvenience. We say the customer that goes alone as the customer served by a *non-sharing service* (NS service). In that case, this impatient customer will be transferred to the NS service, and the number of ℓ -customers waiting for RS service j , $N_{\ell,j}(t)$, decrements by one. We consider a case where the waiting time, for which an ℓ -customer becomes annoyed when waiting for RS service j , is exponentially distributed with rate $\mu_{\ell,j}$, $j \in [J]$ and $\ell \in \mathcal{L}(j)$, and the customers of type $\ell \in [L]$ keep arriving in a Poisson process with rate λ_ℓ .

Let $\mathbf{N}(t) := (N_{\ell,j}(t) : j \in [J], \ell \in \mathcal{L}(j))$ represent the state of counting process at time t , and let $\mathcal{N} := \prod_{j \in [J]} \{0, 1, \dots, N\}^{|\mathcal{L}(j)|}$ represent the state space of counting process $\{\mathbf{N}(t), t \geq 0\}$, where \prod is the Cartesian product and $N \in \mathbb{N}_+$ is a constant representing the maximal number of customers of the same type waiting for a service.

At time $t \geq 0$, if an ℓ_1 and an ℓ_2 -customer depart together with a shared vehicle for some RS service $j \in [J]$ (that is, they are served by RS service j), the system earns an expected reward, $R_j \in \mathbb{R}_+$. Also, an expected reward $\bar{R}_\ell \in \mathbb{R}_+$, $\ell \in [L]$, will be earned by serving an ℓ -customer with the NS service. We assume that for any $j \in [J]$ and $\ell_1, \ell_2 \in \mathcal{L}(j)$, $\bar{R}_{\ell_1} + \bar{R}_{\ell_2} \leq R_j$; the system always earns more or costs less by serving customers with RS services. Also, as mentioned in Section 2.1, customers waiting for a RS service are possibly getting frustrated. Whenever an ℓ -customer becomes annoyed and gets re-assigned from RS service $j \in \mathcal{F}_\ell$ to an NS service, an expected penalty $C_{\ell,j} \in \mathbb{R}_0$ incurs. For notational convenience, when such an expected penalty incurs, we say that an expected reward $-C_{\ell,j}$ is generated.

2.2 A Stochastic Optimization Problem

Consider state-dependent action variables $a_{\ell,j}(\mathbf{N}(t)) \in \{0, 1\}$, $t \geq 0$: if service $j \in [J]$ is selected to accommodate a newly-arrived ℓ -customer at time t , then $a_{\ell,j}(\mathbf{N}(t)) = 1$; otherwise, $a_{\ell,j}(\mathbf{N}(t)) = 0$. Since an ℓ -customer can only be served by either RS services in \mathcal{F}_ℓ , we define $a_{\ell,j}(\cdot) \equiv 0$ for all $j \notin \mathcal{F}_\ell$. These action variables should satisfy

$$\sum_{j \in \mathcal{F}_\ell} a_{\ell,j}(\mathbf{n}) = 1, \quad \forall \mathbf{n} \in \mathcal{N}, \ell \in [L], \quad (1)$$

which guarantees exactly one RS service is selected for each customer type $\ell \in [L]$ when the system is in any state $\mathbf{N}(t) = \mathbf{n} \in \mathcal{N}$. Let $\mathbf{a}(\mathbf{n}) := (a_{\ell,j}(\mathbf{n}) : j \in [J], \ell \in \mathcal{L}(j))$, $\mathbf{n} \in \mathcal{N}$.

For clarity, we refer to the states $\mathbf{n} \in \mathcal{N}$ with $a_{\ell,j}(\mathbf{n}) \equiv 0$ ($\ell \in [L]$, $j \in \mathcal{F}_\ell$) as the *uncontrollable states* for ℓ -customers in RS service j ; precisely, define, for $\ell \in [L]$, $j \in \mathcal{F}_\ell$, $\mathcal{N}_{\ell,j}^{\{0\}} := \{\mathbf{n} \in \mathcal{N} \mid n_{\ell,j} = N\}$,

if $|\mathcal{L}(j)| = 2$; $\{\mathbf{n} \in \mathcal{N} \mid n_{\ell,j} \geq 2\}$, otherwise. The set $\mathcal{N}_{\ell,j}^{\{0\}}$ includes all the uncontrollable states for ℓ -customers in RS service j . The superscript $\{0\}$ indicates the possible value of the action variable $a_{\ell,j}(\mathbf{n})$ for states $\mathbf{n} \in \mathcal{N}_{\ell,j}^{\{0\}}$ is only 0.

These action variables $\mathbf{a}(\mathbf{n})$ for all $\mathbf{n} \in \mathcal{N}$ determine a scheduling policy ϕ for process $\{\mathbf{N}(t), t \geq 0\}$. Let Φ represent the set of all such policies subject to (1). To highlight the relationship between the stochastic process and its underlying policy, we add superscript $\phi \in \Phi$ to its state and action variables; that is, by slightly abusing notation, we rewrite $\mathbf{N}(t)$ and $\mathbf{a}(\cdot)$ as $\mathbf{N}^\phi(t)$ and $\mathbf{a}^\phi(\cdot)$, respectively.

We aim to maximize the long-run average total reward of above described stochastic process $\{\mathbf{N}^\phi(t), \mathbf{a}^\phi(\mathbf{N}^\phi(t)), t \geq 0\}$, where the total reward is the sum of the rewards generated by all RS and NS services:

$$\max_{\phi \in \Phi} \lim_{T \rightarrow +\infty} \frac{1}{T} \mathbb{E} \int_0^T \sum_{j \in [J]} R_j(\mathbf{N}_j^\phi(t), \mathbf{a}_j^\phi(\mathbf{N}^\phi(t))) dt, \quad (2)$$

subject to (1). We refer to this problem described by (2) and (1) as the *ride-matching problem* (RMP).

The RMP is complicated by the large state-space size which increases exponentially in the number of RS services J . For large J , optimal solutions become computationally intractable with conventional techniques, such as value iteration, and we thus resort to simple, heuristic scheduling policies that approximate the optimality.

3 BIVARIATE INDEX POLICY

Consider a special case in our RMP with $|\mathcal{J}_\ell| > 2$ for only one customer type $\ell \in [L]$ and $|\mathcal{J}_{\ell'}| \leq 2$ for all the other $\ell' \in [L] \setminus \{\ell\}$. Also, in this special case, the reward of the self-shared service of any customer type ℓ' in $[L] \setminus \{\ell^*\}$ is always zero; and the reward for the non-self-shared service, if there is one, is sufficiently large so that all ℓ' -customers will be firstly assigned to the non-self-shared service if it is available. The RMP in this special case reduces to $\max_{\phi \in \Phi} \lim_{T \rightarrow +\infty} \frac{1}{T} \mathbb{E} \int_0^T \sum_{j \in \mathcal{J}_{\ell^*}} R_j(\mathbf{N}_j^\phi(t), \mathbf{a}_j^\phi(\mathbf{N}^\phi(t))) dt$ subject to $\sum_{j \in \mathcal{J}_{\ell^*}} a_{\ell^*,j}(\mathbf{N}^\phi(t)) = 1, t > 0$, where, for $\mathbf{n} \notin \mathcal{N}_{\ell^*,j}^{\{0\}}, a_{\ell^*,j}^\phi(\mathbf{n}) \equiv 1$. This problem is a standard RMABP that includes $|\mathcal{J}_{\ell^*}|$ parallel sub-processes, $\{\mathbf{N}_j^\phi(t), \mathbf{a}_j^\phi(\mathbf{N}^\phi(t)), t \geq 0\}$ ($j \in \mathcal{J}_{\ell^*}$), each of which is referred to as a *bandit process* and is all the time controlled by binary actions: selected or not (by an incoming ℓ^* -customer). We use bandit process to indicate the sub-process with binary actions in this very specific example and still refer to generic ones as sub-processes.

For this $j \in \mathcal{J}_{\ell^*}$, let $\ell_1(j) = \ell^*$ where the $\ell_1(j)$ is the first customer type associated with RS service j , and consider a special set of policies, ϕ , determined by the *relaxed* action variables: for $\mathbf{n} \in \tilde{\mathcal{N}}_j, \tilde{\alpha}_j^\phi(\mathbf{n}) := \lim_{t \rightarrow +\infty} \mathbb{E}[\mathbf{a}_j^\phi(\mathbf{N}^\phi(t)) \mid \mathbf{N}_j^\phi(t) = \mathbf{n}]$ which takes values in $\{0, 1\}$. In this very special case, the second element of $\tilde{\alpha}_j^\phi(\cdot)$ has been fixed to be either 0 or 1; precisely, for $\mathbf{n} \in \tilde{\mathcal{N}}_j, \tilde{\alpha}_{j,2}^\phi(\mathbf{n}) \equiv a_2(n_2) = 1$, if $n_2 < N$ and $|\mathcal{L}(j)| = 2$; 0, otherwise. Also, for notational convenience, the expected reward rate of this bandit process is rewritten as $R_j(\mathbf{n}, (a_1, a_2))$ when it is in state $\mathbf{n} \in \tilde{\mathcal{N}}_j$ under a policy ϕ determined by action variable $\tilde{\alpha}_j^\phi(\mathbf{n}) = (a_1, a_2) \in \{0, 1\}^2$.

For $\mathbf{n} \in \tilde{\mathcal{N}}_j \setminus \{(N, 0)\}$, consider a value function of the (η, j) -process, $V_j^\eta(\mathbf{n})$, that solves Bellman equation

$$V_j^\eta(\mathbf{n}) = \max_{a_1 \in \{0,1\}} \left\{ \frac{R_j(\mathbf{n}, (a_1, a_2(n_2))) - a_1 \eta - g}{u_j(\mathbf{n}, a_1)} + \sum_{\mathbf{n}' \in \tilde{\mathcal{N}}_j} p_j^{a_1}(\mathbf{n}, \mathbf{n}') V_j^\eta(\mathbf{n}') \right\}, \quad (3)$$

where $1/u_j(\mathbf{n}, a_1)$ is the expected sojourn time that the process stays in state \mathbf{n} if the employed policy ϕ decides the first element of action variable of the (η, j) -process $\tilde{\alpha}_{j,1}^\phi(\mathbf{n}) = a_1$; $p_j^{a_1}(\mathbf{n}, \mathbf{n}')$ is the probability of the process transitioning from state \mathbf{n} to \mathbf{n}' when $\tilde{\alpha}_{j,1}^\phi(\mathbf{n}) = a_1$ and $\tilde{\alpha}_{j,2}^\phi(\mathbf{n}) = a_2(n_2)$; $\eta \in \mathbb{R}$ and $g \in \mathbb{R}$ are given hyper-parameters. The Bellman equation (3) allows the value function of all states to be computed iteratively, for which, if $g = 0$, the actions selected to maximize the r.h.s. of (3) form an optimal solution ϕ^* that maximizes the expected cumulative reward of the underlying (η, j) -process; that is, whichever $a_1 \in \{0, 1\}$ maximizes the r.h.s. of (3), $\tilde{\alpha}_{j,1}^{\phi^*}(\mathbf{n}) = a_1$. The given real g is an attached criterion for value function $V_j^\eta(\mathbf{n})$, and, from Ross (1992), there exists a specific $g \in \mathbb{R}$ such that the policy ϕ^* formed by actions maximizing the r.h.s. of (3) also maximizes the long-run average reward of the (η, j) -process.

Recall that the Whittle index (if exists) of state $\mathbf{n} \in \tilde{\mathcal{N}}_j \setminus \{(N, 0)\}$ is a specific value of the η that satisfies

$$\frac{R_j(\mathbf{n}, (1, a_2(n_2))) - \eta - g}{u_j(\mathbf{n}, 1)} + \sum_{\mathbf{n}' \in \tilde{\mathcal{N}}_j} p_j^1(\mathbf{n}, \mathbf{n}') V_j^\eta(\mathbf{n}') = \frac{R_j(\mathbf{n}, (0, a_2(n_2))) - g}{u_j(\mathbf{n}, 0)} + \sum_{\mathbf{n}' \in \tilde{\mathcal{N}}_j} p_j^0(\mathbf{n}, \mathbf{n}') V_j^\eta(\mathbf{n}'); \quad (4)$$

the specific η is such that there is no difference of taking either actions at the r.h.s. of (3). Let $\eta_j(\mathbf{n})$ represent the specific η satisfying (4), which, intuitively, represents the marginal reward of taking the action variable to be 1 against 0 in state \mathbf{n} for this bandit process related to RS service j . The Whittle indices can be numerically computed through (relative) *value iteration* methods.

In a generic RMP, for customer type $\ell \in [L]$, the $\eta_j(\mathbf{n})$ ($j \in \mathcal{J}_\ell$, $\mathbf{n} \in \tilde{\mathcal{N}}_j \setminus \tilde{\mathcal{N}}_{\ell,j}^{\{0\}}$) satisfying (4) are still computable along the same lines of the special case by setting $\ell^* = \ell$. Let $\eta_{\ell,j}(\mathbf{n})$ represent the value of such $\eta_j(\mathbf{n})$ for customer type ℓ , and refer to the priorities of state-service (SS) pairs (\mathbf{n}, j) indicated by the descending order of $\eta_{\ell,j}(\mathbf{n})$ as the *Whittle priorities* for customer type ℓ .

In the same vein of the policy derived from the Whittle indices, we propose a policy that always selects the RS service with the highest bivariate index upon an arrival of a customer. We refer to such a policy as the *bivariate index policy* (BI policy). Precisely, consider the stochastic process $\{\mathbf{N}^{\bar{\varphi}}(t), t \geq 0\}$ under BI policy $\bar{\varphi}$ with indices quantified by $\eta_{\ell,j}(\mathbf{n}) \in \mathbb{R}$, $j \in [J]$, $\mathbf{n} \in \tilde{\mathcal{N}}_j \setminus \tilde{\mathcal{N}}_{\ell,j}^{\{0\}}$. For $\ell \in [L]$ and $j \in \mathcal{J}_\ell$, the action variables of BI policy $\bar{\varphi}$ are

$$a_{\ell,j}^{\bar{\varphi}}(\mathbf{N}^{\bar{\varphi}}(t)) = \begin{cases} 1, & \text{if } j \in \min \arg \max_{j' \in \mathcal{J}_\ell} \eta_{\ell,j'}(\mathbf{N}_{j'}^{\bar{\varphi}}(t)), \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

If $\arg \max$ returns more than one number, we break the tie by choosing the smallest. We only care about the priorities of SS pairs indicated by quantified bivariate indices; they can be quantified as infinite values because only relative values of these indices affect the SS-pair priority.

4 NUMERICAL RESULTS

Consider a transportation network with one travel source and multiple destinations uniformly deployed in a coordinate. Each destination is a random variable taking values in a hexagon area, of which the travel distance from the travel source (at position $(0, 0)$) is approximated by the length between the source and its center point. We refer to this distance as the *distance to the destination*. The distance to a destination approximates the expected length of the travel route defined as a random variable in Section 2. In this section, the ride charge of going to a destination is specified to be linear to its distance with slope coefficient $b > 0$ (measured in dollars per unit distance) if a customer takes the NS service; and the charge of a vehicle per unit distance becomes γb , $\gamma \in (1, 2)$, if two customers take a RS service and evenly share the charge of the entire journey. For whichever type, all customers are fairly treated with the same b and γ . In this context, customers are actually classified according to their destinations with one-to-one correspondence between their customer types and destinations. Here, we consider in total 105 destinations and their corresponding customer types labeled by $\ell = 1, 2, \dots, 105$ ($L = 105$).

In this section, we specify the *availability* for two customers of sharing a vehicle by checking the ride charges of corresponding sharing service: if both can save costs through a RS service compared to their NS services, the customers are available of sharing a vehicle with each other; otherwise, not available. Let d_1 and d_2 represent the expected travel distances for the first and second customers, respectively, and $d_{1,2}$ represent the expected travel distance between the destinations of the two customers. Assuming the first customer drops off first, then $d = d_1 + d_{1,2}$ represents the expected total travel distance for the two customers if they share a vehicle. When the sum of expected charge for both customers going alone is larger than the expected total distance of going together, that is, $b(d_1 + d_2) > \gamma b d$, the two customers are available for sharing a vehicle with each other. Note that, with this setting, the drop-off order matters. In particular, we choose $b = 1$ and $\gamma = 1.7$. The expected reward of serving two customers (together) R_j is set to be $2Q - \gamma b(d_{\ell_1} + d_{\ell_1, \ell_2})$ with a normalizing constant $Q = 40$ so that the expected reward rates R_j and \bar{R}_ℓ are guaranteed to be positive. Similarly, let the expected reward for NS service of type ℓ be specified by $\bar{R}_\ell = Q - b d_\ell$, where d_ℓ is the travel distance for an ℓ -customer going alone. In this context, shorter travel routes imply lower travel charges but higher reward rates, and we generate in total 698 RS services ($J = 698$).

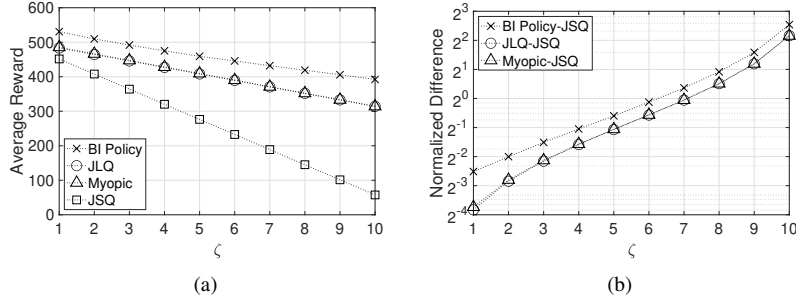


Figure 1. Average reward gained under different policies against penalty parameter ζ : (a) Average reward; and (b) the relative difference of specified policy to JSQ with respect to the average reward.

In addition, we consider identical arrival rates of all customer types, $\lambda_\ell = 0.3$ (customers per minute) for $\ell \in [L]$; while, the renege rate of customer type ℓ waiting for RS service j , $\mu_{\ell,j}$, is assumed to be negatively correlated to the reward R_j and the destination distance d_ℓ . Recall that the inter-arrival times of ℓ -customers (defined in Section 2.1) are exponentially distributed random variables with rate λ_ℓ ; and, similarly, the renege rate $\mu_{\ell,j}$ has also been defined as a parameter of a random variable, intuitively representing the speed at which a waiting customer becomes impatient. Both λ_ℓ and $\mu_{\ell,j} \in (0, 1)$ for $\ell \in [L]$, $j \in \mathcal{J}_\ell$ are measured in number of customers per unit time (per minute). In particular, let $\mu_{\ell,j} = \exp(-vR_j - \beta d_\ell)$ with correlation coefficients $v, \beta \in \mathbb{R}_+$. Plugging in R_j as a function of d_ℓ , we consider $\beta = 0.09$ and $v = 0.03$, which satisfies $\beta > v\gamma b$, to guarantee the negative correlation between $\mu_{\ell,j}$ and d_ℓ . Recall that, once a customer of type ℓ becomes impatient and leaves RS service j by taking its NS service, an instantaneous penalty, $C_{\ell,j}$, incurs. Here, specify $C_{\ell,j} = -\zeta \ln \mu_{\ell,j}$ with $\zeta \in \mathbb{R}_+$. As described in Section 2, we implement a finite maximal number of customers waiting for each RS service, $N = 5$, which avoids explosion of the counting process when the arrival rates of two types of customers differ too much.

In Figure 1, we investigate the (long-run) average reward of the system under three policies against $\zeta = 1, 2, \dots, 10$, where the 95% confidence intervals based on the Student t-distribution of the simulated data are maintained within $\pm 3\%$ of the observed mean. The average reward is the objective of the RMP described in (2), acting as a metric that balances system rewards and service quality. Apart from the BI policy, we consider three more baseline policies: Join-the-Shortest-Queue (JSQ), Join-in-the-Longest-Queue (JLQ), and a policy that myopically assigns customers to the RS services with the highest immediate expected return. JSQ is a simple policy that always assigns a newly-arrived ℓ -customer to the RS service with the least number of waiting ℓ -customers. It is a classical policy that minimizes the average delay of system with parallel birth-and-death processes under mild conditions (see Winston (1977)). On the other hand, for a RS service associated with customer types ℓ_1, ℓ_2 , if the number of waiting ℓ_1 -customers is very large, then more ℓ_2 -customers should be sent here to avoid renege penalties. The JLQ policy is proposed along this idea, which assigns a newly-arrived customer to the RS service with the largest number of customers waiting to share a ride with the new customer. We refer to the baseline policy that myopically assigns customers as the *myopic* policy in the sequel. When the myopic policy is employed and an ℓ -customer arrives at time t , the customer will be assigned to a RS service j with the highest immediate expected return. The immediate expected return is considered as the difference between the immediate expected rewards gained by selecting and not selecting this RS service; that is, $R(\mathbf{N}_j^{\text{myopic}}(t), (1, 1)) - R(\mathbf{N}_j^{\text{myopic}}(t), (0, 1))$ if $\ell_1(j) = \ell$; otherwise, $R(\mathbf{N}_j^{\text{myopic}}(t), (1, 1)) - R(\mathbf{N}_j^{\text{myopic}}(t), (1, 0))$. Here, the action variable associate to the ℓ -customer is set to 1 and 0 indicating that this RS service is and is not selected, respectively. Tie cases are broken randomly.

In Figure 1(a), the BI policy achieves clear advantages against JSQ, JLQ and the myopic policy with respect to the average reward regardless of the choice of ζ . We remark that the average reward demonstrates an approximate linear relation against the penalty term ζ . This is mainly because the customer demand is uniform and the first derivative of the proportion of renege customers does not change much. In particular, the performance of JSQ degrades much faster than the others as the penalty parameter ζ increases. Figure 1(b) demonstrates the relative differences of BI, JLQ and the myopic policy to JSQ with respect to the average reward as ζ increases from 1 to 10. Precisely, the relative difference of policy ϕ_1 to policy ϕ_2 is $(\rho^{\phi_1} - \rho^{\phi_2})/\rho^{\phi_2}$, where ρ^ϕ represents the average reward under policy ϕ . In Figure 1(b), the relative differences of BI, JLQ and the myopic policy to JSQ increases almost exponentially in ζ . The bivariate priorities of SS pairs for the BI policy and the priorities of RS services for the myopic policy will both be automatically modified when system parameters change, and JLQ is a policy aiming at avoiding potential renege penalties. While,

JSQ does not change its schedules of customers for different ζ values, and thus its performance degrades significantly. In Figure 1, although the deviation between curves for the myopic policy, JLQ and BI is always much smaller than that between JSQ and BI, BI outperforms the myopic policy and JLQ by above 9% of average reward for all the tested ζ and the relative difference of BI to JLQ (or the myopic policy) reaches around 24% for $\zeta = 10$.

5 CONCLUSIONS

We have proposed a scalable BI policy for the RMP, which is complicated by the heterogeneity of passengers who are able to share vehicles with each other, the renegeing passengers and the large size of the problem. The effectiveness of the BI policy has been further demonstrated through extensive numerical simulation with real-world settings. Compared to the baseline policies, BI achieves over 20% higher average rewards. To the best of our knowledge, no existing work has studied the multi-rider matching problem with enabled renegeing behaviors that has been analyzed in this paper. This is also the first attempt of applying the Whittle relaxation technique to an extended RMABP where sub-processes with multiple action variables subject to different action constraints are inevitable.

REFERENCES

- Agatz, N., Erera, A., Savelsbergh, M. & Wang, X. (2012), 'Optimization for dynamic ride-sharing: A review', *European Journal of Operations Research* **223**(2), 295–303.
- Clarke, M. & Freilich, J. (1987), Airport shared-ride taxi programs in new york, chicago and boston, Technical report, Department of Transportation, US.
- Dong, Y., Wang, S., Li, L. & Zhang, Z. (2018), 'An empirical study on travel patterns of internet based ride-sharing', *Transportation Research Part C: Emerging Technologies* **86**, 1–22.
- Hong, Z., Chen, Y., Mahmassani, H. S. & Xu, S. (2017), 'Commuter ride-sharing using topology-based vehicle trajectory clustering: Methodology, application and impact evaluation', *Transportation Research Part C: Emerging Technologies* **85**, 573–590.
- Lee, A. & Savelsbergh, M. (2015), 'Dynamic ridesharing: Is there a role for dedicated drivers?', *Transportation Research Part B: Methodological* **81**, 483–497.
- Lee, D.-H., Wang, H., Cheu, R. L. & Teo, S. H. (2004), 'Taxi dispatch system based on current demands and real-time traffic conditions', *Journal of Transportation Research Record* **1882**, 193–200.
- Lee, K.-T., Lin, D.-J. & Wu, P.-J. (2005), 'Planning and design of a taxipooling dispatching system', *Transportation research record* **1903**(1), 86–95.
- Lee, K. T., Wu, P. J. & Wang, S. H. (2004), The planning and design of taxipooling on feeder system, in 'Proc. IEEE International Conference on Networking, Sensing and Control', Taiwan, pp. 376–381.
- Levin, M. W., Kockelman, K. M., Boyles, S. D. & Li, T. (2017), 'A general framework for modeling shared autonomous vehicles with dynamic network-loading and dynamic ride-sharing application', *Computers, Environment and Urban Systems* **64**, 373–383.
- Liu, Z., Wang, S., Huang, K., Chen, J. & Fu, Y. (2018), 'Practical taxi sharing schemes at large transport terminals', *Transportmetrica B: Transport Dynamics* **7**(1), 596–616.
- Lokhandwala, M. & Cai, H. (2018), 'Dynamic ride sharing using traditional taxis and shared autonomous taxis: A case study of NYC', *Transportation Research Part C: Emerging Technologies* **97**, 45–60.
- Lyu, Y., Lee, V. C., Ng, J. K.-Y., Lim, B., Liu, K. & Chen, C. (2019), 'Flexi-sharing: A flexible and personalized taxi-sharing system', *IEEE Transactions on Vehicular Technology* **68**(10), 9399–9413.
- Papadimitriou, C. H. & Tsitsiklis, J. N. (1999), 'The complexity of optimal queuing network control', *Mathematics of Operations Research* **24**(2), 293–305.
- Ross, S. M. (1992), *Applied probability models with optimization applications*, Dover Publications (New York).
- Wang, H. & Yang, H. (2019), 'Ridesourcing systems: A framework and review', *Transportation Research Part B: Methodological* **129**, 122–155.
- Whittle, P. (1988), 'Restless bandits: Activity allocation in a changing world', *Journal of Applied Probability* **25**, 287–298.
- Winston, W. (1977), 'Optimality of the shortest line discipline', *Journal of Applied Probability* **14**(1), 181–189.
- Zhong, M. & Liu, Y. (2020), Research on the Evaluation Index System of the Taxi Operation and Service in Airport Hubs, in 'in Proceeding of 8th International Conference on Traffic and Logistic Engineering', Shenzhen, China.