A Privacy-Preserving Order Dispatch Scheme for Ride-Hailing Services

Yubin Duan†, Guoju Gao†‡, Mingjun Xiao‡, and Jie Wu†
†Department of Computer and Information Sciences, Temple University, USA
‡School of Computer Science and Technology, University of Science and Technology of China, P. R. China
Email: yubin.duan@temple.edu, gaoguoju@mail.ustc.edu.cn, xiaomj@ustc.edu.cn, jiewu@temple.edu

***Abstract*—The ride-hailing system has become popular around the world. The Service Providers (SPs) such as Uber and Didi dispatch passenger orders based on their location information. However, one concern from the public is whether the SPs could protect the location privacy of passengers. In this paper, we propose an order dispatch scheme that could preserve the location privacy of passengers based on their requirements. Our scheme uses cloaking regions in which the SPs cannot distinguish actual locations of passengers. The trade-off is the loss of matching performance or social welfare, i.e., the increase in the overall pick-up distance. We formulate the problem as maximizing the social welfare (or minimizing the overall pick-up distances) under privacy requirements of passengers. A bipartite-matching-based scheme is investigated, and we provide a theoretical bound on the matching performance under specific privacy requirements. Nevertheless, minimizing the overall pick-up distances does not consider the interest of each individual passenger. Passengers with low privacy requirements may be matched with drivers far from them. Therefore, we further propose a pricing scheme that could make up for the individual loss by allocating discounts on their riding fares. Especially, three discount allocation strategies are proposed in this paper. Experiments on both real-world and synthetic datasets show the efficiency of our scheme.**

***Index Terms*—order dispatch, pricing, privacy, ride-hailing**

I. Introduction

Ride-hailing service has rapidly developed nowadays [1, 2]. Service Providers (SPs) such as Uber or Didi already have a large number of registered users. Although the ride-hailing services could bring travel convenience to users, they also carry privacy risks. Specifically, the SP could easily gather millions of travel traces per day. By digging in these travel records, the SP could collect a large amount of private information of passengers. For example, the SP could infer living or working locations of passengers, or even their habits and interests [3, 4]. Exposing such information to unauthorized organizations may bring location-based scams to passengers. It may further cause economic or social reputation damage to passengers. Harmful consequences of privacy leakage in ride-hailing systems has been reported in [5]. Therefore, there is a strong need to protect the location privacy of passengers in the ride-hailing systems.

Existing researches on ride-hailing are mainly based on spatial cloaking [6, 7] and/or built on homomorphic encryption [8–10]. In the spatial cloaking approaches, passengers report cloaking regions to the SP and their actual locations are indistinguishable in these regions. To match passengers with drivers, existing researches propose to let each passenger choose the nearest available driver. The SP sends locations of nearby drivers to passengers without knowing their actual locations. However, this approach does not consider the social welfare. That is, the overall pick-up distance is not optimized. As a result, the efficiency of the ride-hailing system is not fully used. In appraoches based on the homomorphic encryption, the SP could attain some non-sensitive information, such as the distance between a passenger's source location and a driver's location. With such information, it is hard for the SP to infer the actual locations of passengers or drivers. However, homomorphic encryption requires additional communication overheads between drivers and passengers, which brings difficulties for its real-world implementation.

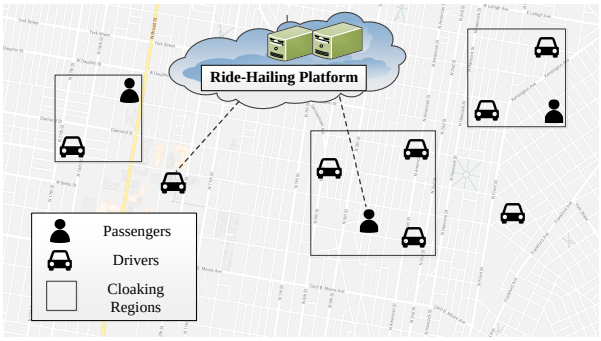


Fig. 1. The application scenario.

An attacking model toward the simple spatial cloaking approach is proposed in [11]. They assume the SP is honest but curious. It means the SP would send accurate driver locations to passengers but try to infer passenger locations. The Voronoi diagram [12] is used in their attacking model, since the SP knows that each passenger would choose the nearest available driver. Although an enhanced dispatch scheme is proposed, the social welfare is not considered in their work.

In this paper, we propose to optimize the social welfare (or to minimize the overall pick-up distance) while ensuring privacy requirements of passengers and maintaining a low communication overhead. A privacy-preserving order dispatch scheme based on spatial cloaking is introduced. The difference is that our scheme lets the SP globally match passengers with drivers based on pick-up distances in a centralized manner. In this way, each passenger may no longer be matched with the nearest available driver. Consequently, the SP cannot infer the passenger locations by using the attacking model in [11]. The trade-off is that the performance of matching is affected, since the SP only knows cloaking regions rather than actual locations of passengers. To best of our knowledge, our paper is the first one that investigates this trade-off and provides a theoretical bound on the matching performance under given privacy requirements.

On the other hand, minimizing the overall pick-up distance does not consider the interest of each individual. For example, passengers with low privacy requirements may be matched with drivers who are far from them. We propose to make up the loss of each individual, which has not been fully discussed in previous researches. A pricing scheme is introduced. Specifically, the SP would first collect additional fees from passengers for their privacy requirements, since the performance of the matching is affected by these requirements. The additional fee is positively correlated with the privacy requirement. Then, the SP would allocate part of the collected fees as discounts to make up individual losses. It is challenging to determine a closed form equation to describe the relation between privacy requirements and their side effects on matching performance. The reason is that the performance of global matching is not only determined by each passenger's privacy requirement, but also affected by other passengers' settings. In this paper, three discount allocation strategies are investigated.

An application scenario of our scheme is shown in Fig. 1. Each passenger would send a cloaking region which contains his/her actual location to the SP. The size of the cloaking region is chosen by passengers based on their privacy requirements. Instead of letting passengers choose drivers, the SP would globally match drivers with passengers based on locations of cloaking region centers and drivers. By using this scheme, the SP could not infer actual locations of passengers by using the Voronoi diagram. However, the trade-off is that passengers cannot be matched with the optimal drivers in terms of the social welfare or their own interests, since their actual locations are unknown to the SP. After passengers report their satisfaction, the SP could allocate discounts based on our pricing scheme. Finally, passengers could contact the assigned drivers on secure channels and start the riding.

The contributions of this paper are summarized as follows:

- We propose a privacy-preserving ride-hailing scheme based on the global matching, in which passenger locations cannot be inferred by the attacking model in [11] and no significant communication overhead is introduced.
- We evaluate the performance loss brought by using inaccurate passenger locations in matching. A theoretical bound on the loss is given.
- We investigate three discount strategies that could make up for the performance loss in the matching process.

The remainder of the paper is organized as follows. Section II reviews related works. Section III presents our system model. Section IV shows the ride matching algorithm and analyzes its performance and privacy-preserving properties. Section V introduces our discount allocation strategies. Section VI simulates our approaches on both synthetic and real-world datasets. Finally, Section VII concludes the paper.

II. Related Work

A. Location Privacy in Ride-Hailing System

Researches on the privacy in ride-hailing systems [11, 13–18] mainly have two different approaches. One approach is based on location cloaking [6, 7, 11]. The other approach is based on homomorphic encryption [8–10]. In the location cloaking approach, instead of uploading their actual locations to the SP, passengers would report cloaking regions centered at arbitrary fake locations within their nearby areas. Their actual locations are not distinguishable within cloaking regions [19]. The SP would send locations of all available drivers in cloaking regions to passengers. Then, passengers can choose drivers based on some metrics. [19] propose to let each passenger choose the nearest driver. However, [11] points out that the SP could infer actual passenger locations to a certain degree by using Voronoi diagram [12]. To enhance the privacy level, [11] proposes to choose relatively nearer drivers with higher probability. Although the possibility of inferring actual locations of passengers is decreased, the social welfare is not considered. In this paper, besides caring about the privacy of each passenger, we also aim to maximize the social welfare with certain theoretical bound, and a global matching based scheme is proposed. Although [2] proposes to optimize the social welfare, the privacy issue is not considered.

B. Discount Allocation Problems

To the best of our knowledge, there is little research work on the discount allocation algorithm (also called pricing for privacy) for the ride-hailing systems. The authors in [20] design a usage-based dynamic pricing scheme with privacy preservation for smart grid, in which they enable the electricity price to correspond to the electricity usage in real time. Zhuo *et al.* in [21] study the tradeoff between the amount of traffic being offloaded and the users' satisfaction in 3G network, and further propose a novel incentive framework to motivate users to leverage their delay tolerance for 3G traffic offloading. Essentially, the discount allocation algorithm is used to motivate individuals to participate in the privacy-preserving ride-hailing system by providing them some benefits (i.e., discount).

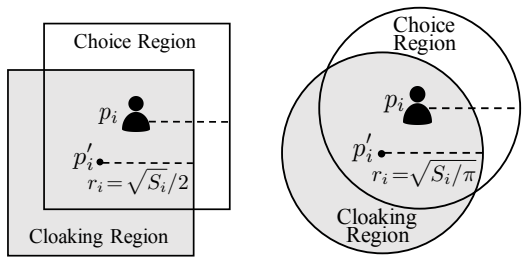
The most common incentive mechanism is the auction model [22], such as generalized second-price auction [23], Vickrey-Clarke-Groves (VCG) auction [24], etc. The VCG auction is a type of sealed-bid auction of multiple items, in which bidders submit bids that report their valuations for the items, without knowing the bids of the other bidders. Then, the auction system assigns the items in a socially optimal manner: it charges each individual the harm they cause to other bidders. It gives bidders an incentive to bid their true valuation, by ensuring that the optimal strategy for each bidder is to bid their true valuation of the items. In this paper, we adopt the idea of payment determination in VCG auction while taking the fairness of discount into consideration.

III. Model

A. Preliminaries

We first introduce the notations used in the paper. Let $\mathcal{P}$ denote the set of passenger actual locations, and a passenger location in the set is denoted as p_i, $1 \leq i \leq |\mathcal{P}|$. Let $\mathcal{D}$ denote the set of driver locations, and each driver location is denoted as $d_j, 1 \leq j \leq |\mathcal{D}|$.

To protect the location privacy, each passenger would construct a cloaking region based on his/her privacy requirement.



(a) Square (b) Circle

Fig. 2. The illustration of the cloaking region.

The cloaking region [25] of a passenger is a geographical region in which the passenger's actual location is indistinguishable from other points within the region. The privacy of each passenger is quantified by the area of the cloaking region. By default, we assume that the shape of each cloaking region is square as shown in Fig. 2(a). Actually, its shape could also be circle as shown in Fig. 2(b), and the conclusions of our paper can be easily extended. Formally, let R_i and S_i denote the cloaking region and the privacy requirement of passenger p_i, respectively. To generate the cloaking region R_i, the passenger p_i would randomly chose a location p_i' as the center of R_i. To ensure that the actual location p_i is contained in the cloaking region, p_i' should be chosen from the choice region as shown in Fig. 2. The area of the cloaking region R_i is S_i, or the side length of R_i is $2r_i = \sqrt{S_i}$ for square cloaking regions. The choice region has the same size as the cloaking region. After generating the cloaking region R_i, the passenger at p_i would send R_i instead of the actual location to the SP.

Due to the heterogeneous privacy requirements of passengers, the performances of global matching would deteriorate. For this reason, the passengers are required to pay additionally for their riding fees. Let T denote the total additional payment by passengers. Since both passengers and drivers may suffer loss in the global matching process, the system should share the profits T with them. The detailed allocation strategy among individuals would be studied in Section V.

The performance of the ride matching scheme is evaluated by the social welfare (or the overall pickup distance). Formally, let $dis(\cdot,\cdot) : \mathbb{R}^2 \mapsto \mathbb{R}$ denote the distance function. The pickup distance for the passenger at p_i is $dis(p_i, d_j)$ where d_j is the driver matched with passenger p_i. Correspondingly, the social welfare is defined as $W = -\sum_{p_i \in \mathcal{P}} dis(p_i, d_j)$, which is the negation of the overall pickup distance. It is because a longer pickup distance corresponds to lower social welfare.

B. Problem formulation

In this paper, we aim to maximize the social welfare while guaranteeing the privacy requirements of passengers. Formally, our problem is defined as following:

$$\max \quad W = -x_{ij} dis(p_i, d_j) \tag{1}$$

$$\text{s.t.} \quad \sum_{1 \le j \le |\mathcal{D}|} x_{ij} = 1, \forall p_i \in \mathcal{P} \tag{2}$$

$$\|p_i - p_i'\|_\infty \le \sqrt{S_i}/2, \forall p_i \in \mathcal{P} \tag{3}$$

$$x_{ij} \in \{0,1\}, \forall p_i \in \mathcal{P}, \forall d_j \in \mathcal{D} \tag{4}$$

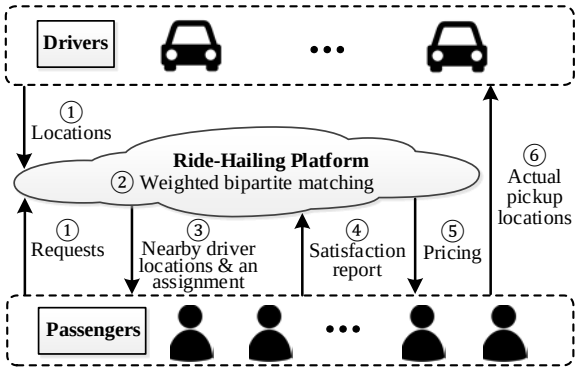


Fig. 3. The system framework.

where x_{ij} is the decision variable. $x_{ij} = 1$ if and only if the passenger p_i is matched with the driver d_j. Eq. (1) represents our objective which is to maximize social welfare, while the value of p_i' instead of p_i is known. Eq. (2) is the matching constraint that means each passenger should be paired with one driver. Eq. (3) is the privacy constraint, which means that actual location of each passenger p_i should be contained in the cloaking regions. $\|\cdot\|_\infty$ denotes the l_∞-norm. Eq. (3) represents that the p_i' would be located within the square region centered at p_i with side length $\sqrt{S_i}$. Eq. (4) is the binary constraint for the decision variable.

IV. Privacy-preserving dispatch process

A. System Framework

The framework of our dispatch system is shown in Fig. 3. In the first step, drivers need to upload their locations to the SP, and passengers need to send their cloaking regions to the SP. The privacy requirements of passengers is represented by the sizes of cloaking regions. Then, in the second step, the SP would match passengers with drivers by using the weighted bipartite matching algorithm [2]. Details of the matching process are introduced in Section IV-B. This step aims to maximize social welfare or to minimize the overall pick-up distance. Then in the third step, the SP would send the matching results to passengers along with locations of drivers around the cloaking regions. The SP broadcasts driver locations with the purpose of letting passengers evaluate their satisfaction with the matching results. In the fourth step, passengers would report their satisfactions to the SP and the SP applies the discount allocation strategies (which is introduced in Section V) to make up for the individual loss in the fifth step. After receiving the discount, each passenger could contact the matched driver in a private communication channel and share his/her actual location to the driver. Procedures of our order dispatch scheme is shown in Algorithm 1.

B. Matching with Cloaking Regions

In our scheme, to protect the location privacy of passengers in the ride-hailing systems, they upload cloaking regions instead of their precise locations to the SP. In our matching process, we propose to use centers of the cloaking regions as obfuscated locations of passengers. Then, we apply the weighted bipartite matching algorithm with obfuscated

Algorithm 1 The order dispatch scheme

Input: Passenger reported locations $\mathcal{P}'$, driver locations $\mathcal{D}$
Output: Dispatch results for passengers

1: Construct a weighted bipartite matching graph $G = (V, E)$. $V = \mathcal{P}' \bigcup \mathcal{D}$, $E = -dis(\mathcal{P}' \times \mathcal{D})$.
2: $M \leftarrow$ weighted bipartite matching on G.
3: send corresponding matching result M to each passengers along with locations of drivers near the cloaking region.
4: receive satisfactions from passengers.
5: Pricing for passengers $\leftarrow$ Discount Allocation Algorithm.
6: **return** M

locations of passengers and driver locations. Although the matching result is not optimal with respect to passenger actual locations, we show that there is a theoretical upper bound for its difference from the optimal value.

In the matching process, the SP first constructs a bipartite matching graph. Specifically, the matching graph $G = (V, E)$, where $V = \mathcal{P}' \cup \mathcal{D}$ and $E = -dis(\mathcal{P}' \times \mathcal{D})$. It means that the matching graph is bipartite. One side contains elements in set $\mathcal{P}'$ and the other side contains elements in set $\mathcal{D}$. Weight of the edge between a $p' \in \mathcal{P}'$ and a $d \in \mathcal{D}$ is the negation of the geographical distance between the obfuscated location p' and driver's location d. We use the negation of the distance as the edge weight, since the social welfare decreases if the total distances increase. Our objective is to maximize the social welfare, which is equivalent to maximizing the negative of sum distance between passengers and drivers.

An example of our dispatch scheme is shown in Fig. 4 and 5. In Fig. 4, the blue solid line represents the optimal weighed bipartite matching founded by our scheme and the red dashed line represents the optimal weighted bipartite matching between passenger actual locations and drivers. The distance between the locations used in the example is given in Fig. 5. In this example, we can find out that the overall pick-up distance of matching with obfuscated locations is not optimal since the actual passenger locations are unknown in the matching process. Specifically, passengers p_1, p_2 and p_3 should be matched with the drivers d_1, d_2 and d_3 respectively if their actual locations are used in the matching process. The optimal overall pick-up distance is $dis(p_1, d_1)+dis(p_2, d_2)+dis(p_3, d_3) = 2\sqrt{2}+1+2 = 5.83$. However, p_1, p_2 and p_3 are matched with d_3, d_1 and d_2 respectively by applying our ride matching scheme. The corresponding overall pick-up distance is $dis(p_1, d_3)+dis(p_2, d_1)+dis(p_3, d_2) = \sqrt{10}+1+\sqrt{5} = 7.40$. Note that we should use the actual locations rather than obfuscated locations when calculating the overall pick-up distance of our scheme, although the matching is based on obfuscated locations. The reason is that drivers need to pick up passengers at their actual locations instead of obfuscated locations. From the example, we can find out that the overall pick-up distance increases and the social welfare decreases when matching with obfuscated locations. The extra pick-up distances for drivers are wasted, and we show an upper bound of this waste in the following subsection.

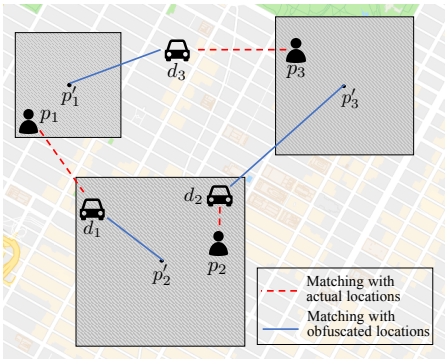


Fig. 4. The effect of cloaking regions in the matching process.

	d_1	d_2	d_3
p_1	$2\sqrt{2}$	$\sqrt{17}$	$\sqrt{10}$
p_2	2	1	$\sqrt{10}$
p_3	$3\sqrt{2}$	$\sqrt{5}$	2

(a) Distances between actual locations and drivers.

	d_1	d_2	d_3
p'_1	$\sqrt{10}$	$\sqrt{13}$	2
p'_2	$\sqrt{2}$	$\sqrt{5}$	4
p'_3	$2\sqrt{5}$	$\sqrt{5}$	$\sqrt{10}$

(b) Distance between reported locations and drivers.

Fig. 5. The distance table of the example.

C. Analysis of Matching Performance and Privacy

We first analyze the performance of our ride matching scheme. Let OPT denote the optimal overall pick-up distance. It can be calculated by using the weighted bipartite matching on actual locations of passengers and locations of drivers. Formally, $OPT = \sum_{p_i \in \mathcal{P}} dis(p_i, d_j)$, where d_j is the driver matched with p_i by using passenger's actual locations in the bipartite matching. In our scheme, the SP could only perform the bipartite matching based on obfuscated locations of passengers. Let M denote the overall distance between obfuscated locations of passengers and locations of drivers matched by our scheme (i.e., based on passengers' obfuscated locations). Formally, $M = \sum_{p_i \in \mathcal{P}} dis(p'_i, d'_j)$, where d'_j is the driver matched with p'_i by applying our scheme. Note that M is not the overall pick-up distance if our scheme is used. The reason is that drivers should pick up passengers at their actual locations rather than obfuscated locations. Let M' denote the overall pick-up distance of our scheme, i.e., the sum of distance between passengers' actual locations p_i and locations of drivers d'_j. Formally, $M' = \sum_{p_i \in \mathcal{P}} dis(p_i, d'_j)$. Fig. 6 illustrates the meaning of these notations.

***Theorem** 1:* Our matching scheme guarantees that $M' < OPT + \sqrt{2}\sum_{p_i \in \mathcal{P}} \sqrt{S_i}$, where S_i is the privacy setting.

Proof: The theorem is proved by using the triangle inequality and the optimal property of the weighted bipartite matching. We illustrate the proof in Fig. 6.

By definition, we have $OPT = \sum_{p_i \in \mathcal{P}} dis(p_i, d_j)$, $M = \sum_{p_i \in \mathcal{P}} dis(p'_i, d'_j)$, and $M' = \sum_{p_i \in \mathcal{P}} dis(p_i, d'_j)$. Since the obfuscated location of each passenger must be located within the cloaking region, we have that $dis(p_i, p'_i) \leq \sqrt{2} r_i$.

Based on the triangle inequality, we have that

$$M' = \sum_{p_i \in \mathcal{P}} dis(p_i, d'_j)$$

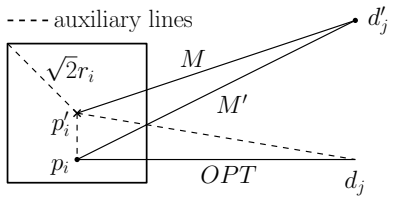


Fig. 6. The illustration of the proof. OPT means matching with actual locations, M represents the matching with obfuscated locations, and M' shows the pick-up distance (picking up at passengers' actual locations).

$$\leq \sum_{p_i \in \mathcal{P}} [dis(p'_i, d'_j) + dis(p_i, p'_i)] = M + \sum_{p_i \in \mathcal{P}} dis(p_i, p'_i).$$

Similarly, $\sum_{p_i \in \mathcal{P}} dis(p'_i, d_j) \leq OPT + \sum_{p_i \in \mathcal{P}} dis(p_i, p'_i)$.

Based on the optimal property of the weighted bipartite matching, we have that $M \leq \sum_{p_i \in \mathcal{P}} dis(p'_i, d_j)$.

Above all, we have

$$\begin{aligned} M' &\leq M + \sum_{p_i \in \mathcal{P}} dis(p_i, p'_i) \leq \sum_{p_i \in \mathcal{P}} dis(p'_i, d_j) + \sum_{p_i \in \mathcal{P}} dis(p_i, p'_i) \\ &\leq OPT + 2\sum_{p_i \in \mathcal{P}} dis(p_i, p'_i) \leq OPT + 2\sum_{p_i \in \mathcal{P}} \sqrt{2} r_i. \end{aligned}$$

Note that $r_i = \sqrt{S_i}/2$, then $M' < OPT + \sqrt{2} \sum_{p_i \in \mathcal{P}} \sqrt{S_i}$. ■

Then, we analyze the privacy related property.

Lemma 1: Our scheme could achieve the strong privacy for actual locations of passengers defined in [11].

Proof: The strong privacy holds since the SP could not infer the passenger locations by knowing the locations of matched drivers. Following the proof in [11], we could conclude that the strong privacy holds. ■

V. Discount Allocation Strategies

As we have mentioned before, the proposed privacy-preserving order-dispatch system concentrates on maximizing the social welfare, i.e., minimizing the total pick-up distances. To achieve this goal, we adopt the optimal matching algorithm during the order-dispatch process. In fact, each individual user in the system is rational and selfish. This means that each passenger always prefers the driver closest to him, and each driver wants to pick up the closest passenger. Although the proposed algorithm can obtain fine global performances, the individual users including passengers and drivers may suffer some losses. Therefore, in order to stimulate individual users to participate in the privacy-preserving order-dispatch system, each participant should receive a certain incentive (called discount in this paper).

We first let T denote the total additional payment by passengers for their privacy requirements, and then introduce the percentage parameters $0 < \gamma, \kappa < 1$. After that, T will be divided into three shares, as shown in Table I. Here, the specific values of γ and κ are determined after three-party (i.e., system, passengers, and drivers) negotiation.

Next, the problem is how to allocate the shared profits (i.e., discount allocation) to each individual. We propose three allocation strategies in the following, that is, individual-loss-based discount allocation strategy, individual-contribution-based discount allocation strategy, and joint discount allocation strategy. Here, since the exact locations of passengers are unknown to drivers, the individual-loss-based and joint discount allocation strategies cannot be applied for drivers.

A. Individual-Loss-Based Discount Allocation

We first design a discount allocation strategy from the perspective of individual loss. Compared to the previous free-choice-based ride-hailing models, our algorithm based on perfect/optimal matching can achieve excellent global performances. However, some individual users may suffer certain losses. These individual users may be unwilling to participate in the system if they cannot receive certain compensations. Note that the total discounts for all passengers are determined, i.e., $\gamma \times T$, and we need to allocate $\gamma \times T$ to each individual passenger efficiently. To this end, we design an individual-loss-based discount strategy for passengers as follows.

In our privacy-preserving order-dispatch system, the locations of drivers are not protected. For the passenger $p_i \in \mathcal{P}$, his/her distance to the nearest driver is known. After p_i is assigned to driver d'_j which is not the nearest driver, the distance between the true location of p_i and d'_j can be calculated. Thus, it is easy to compute the difference (i.e., individual loss) between the actual distance and the nearest distance for p_i, denoted as Δl_i. Note that Δl_i here is a non-negative value, i.e., $\Delta l_i \geq 0$. This is true because the best assignment result for p_i is its nearest driver and now $\Delta l_i = 0$; In other cases, we always have $\Delta l_i > 0$. Moreover, the larger Δl_i indicates the more loss for p_i. Also, the system must compensate more money for the passengers with the larger loss. Based on this observation, we design the discount allocation strategy as follows.

For each passenger $p_i \in \mathcal{P}$, the allocated discount, denoted as t_i, is proportional to its loss Δl_i, i.e.,

$$t_i = \gamma \times T \times \frac{\Delta l_i}{\sum_{p_k \in \mathcal{P}} \Delta l_k}. \tag{5}$$

This intuition-based allocation strategy can ensure the fairness for passengers. Here, we introduce the concept of "fairness". In this system, each passenger $p_i \in \mathcal{P}$ has an additional expense, denoted as Δ_i. When there is no allocated discount, the additional expense is equal to its loss, i.e., $\Delta_i = \Delta l_i$. When each passenger p_i receives a discount t_i, the additional expense is $\Delta_i = \Delta l_i - \rho \cdot t_i$ where ρ denotes the balanced parameter.

Definition 1: Let σ denote the variance of the additional expense $\{\Delta_i | p_i \in \mathcal{P}\}$, that is, $\sigma^2 = \sum_{p_i \in \mathcal{P}} (\Delta_i - \overline{\Delta})^2 / |\mathcal{P}|$, in which $\overline{\Delta} = \sum_{p_i \in \mathcal{P}} \Delta_i / |\mathcal{P}|$ means the average value of the additional expense. Here, the small σ indicates that the additional expense of passengers has little difference. Thus, the smaller σ, the fairer this system.

In the individual-loss-based discount allocation strategy, the passengers with larger loss will receive more discount. Obviously, the variance of $\Delta l_i - \rho \cdot t_i$ is smaller than that of Δl_i. Thus, we can prove that the individual-loss-based discount allocation strategy is fair. Moreover, we will evaluate the metric of fairness in the simulation section to verify the efficiency of the discount allocation strategy.

TABLE I
PROFITS SHARE.

Total Profits	Passengers	Drivers	System
T	$\gamma \times T$	$(1-\gamma)\kappa \times T$	$(1-\gamma)(1-\kappa) \times T$

B. Individual-Contribution-Based Discount Allocation

In fact, the individual-loss-based discount allocation strategy only concerns the individual loss without considering global performances. In other words, the passengers with large individual loss may have little effect on the global performances, while other passengers with a small individual loss might have an important impact. For example, there is such a passenger in this system, whose individual loss is small. If we make discount allocations without involving this passenger, the achieved overall social welfare may be lowered drastically. Therefore, the discount allocated to a passenger depends on not only its individual loss but also its contributions. The contribution of a passenger means the increased social welfare after involving this corresponding passenger. In this subsection, we will introduce the individual-contribution-based discount allocation strategy.

The contribution of a passenger is calculated based on the difference between the social welfare value that includes this passenger and one that excludes this passenger. Actually, the idea of individual-contribution-based discount strategy is similar to the payment determination method used in Vickrey-Clarke-Groves (VCG) auction mechanism [24]. Here, the total profit shared by the system to all passengers is determined as before, i.e., $\gamma \times T$. In our problem, we give each passenger the discount proportional to its contribution. For simplicity, let C_i denote the contribution of $p_i \in \mathcal{P}$. To acquire the value of C_i, we need to compute the original social welfare based on all passengers, which is denoted as W. Then, we let W_{-p_i} denote the social welfare based on the passengers excluding $p_i \in \mathcal{P}$. Note that here $W \geq W_{-p_i}$ for $\forall p_i \in \mathcal{P}$. According to this, we can calculate the contribution of p_i as $C_i = W - W_{-p_i}$. Since the discount of a passenger $p_i \in \mathcal{P}$ is proportional to its contribution, we have:

$$t_i = \frac{(\gamma \times T) \times C_i}{\sum_{p_k \in \mathcal{P}} C_k} = \frac{(\gamma \times T) \times (W - W_{-p_i})}{W \times |\mathcal{P}| - \sum_{p_k \in \mathcal{P}} W_{-p_k}}. \quad (6)$$

Furthermore, the individual-contribution-based discount allocation strategy can also be applied to drivers. As we introduce before, the total profits allocated to all drivers are determined, i.e., $(1-\gamma) \times \kappa \times T$. At the same time, the individual contribution for one driver (i.e., d_j), denoted as G_j, can be calculated as the process for passengers. The original social welfare based on all drivers is denoted as W, and we denote W_{-t_j} as the newly obtained social welfare value after excluding $d_j \in \mathcal{D}$. Similarly, we use $G_j = W - W_{-t_j}$ to denote the contribution of $d_j \in \mathcal{D}$. Since the total profits enjoyed by all drivers are $(1-\gamma) \times \kappa \times T$, we provide each individual driver d_j with the following discount:

$$t_j = \frac{(1-\gamma) \cdot \kappa \cdot T \times G_j}{\sum_{d_x \in \mathcal{D}} G_x} = \frac{(1-\gamma) \cdot \kappa \cdot T \times (W - W_{-d_j})}{W \times |\mathcal{D}| - \sum_{d_x \in \mathcal{D}} W_{-d_x}}. \quad (7)$$

Algorithm 2 Discount Allocation Algorithm for Passengers

Input: $\mathcal{P}$, $\mathcal{D}$, γ, κ, λ, T
Output: t_i for $\forall p_i \in \mathcal{P}$.

1: Initialization: $t_i = 0$ for $\forall p_i \in \mathcal{P}$;
2: Platform computes the total social welfare, i.e., W;
3: **for** $p_i \in \mathcal{P}$ **do**
4: p_i calculates its individual loss, i.e., Δl_i, and then sends the value to the system platform;
5: Platform re-matches passengers $\mathcal{P}/\{p_i\}$ and drivers $\mathcal{D}$, and further computes the new social welfare, i.e., W_{-p_i};
6: Platform computes the contribution, i.e., $C_i = W - W_{-p_i}$;
7: **for** $p_i \in \mathcal{P}$ **do**
8: Platform calculates the discount for p_i, denoted as t_i,

$$t_i = \begin{cases} \frac{\gamma T \Delta l_i}{\sum_{p_k \in \mathcal{P}} \Delta l_k}, & \text{Eq. (5) : individual loss;} \\ \frac{\gamma T C_i}{\sum_{p_k \in \mathcal{P}} C_k}, & \text{Eq. (6) : individual contribution;} \\ \gamma T (\frac{\lambda \Delta l_i}{\sum_{p_k \in \mathcal{P}} \Delta l_k} + \frac{(1-\lambda) C_i}{\sum_{p_k \in \mathcal{P}} C_k}), & \text{Eq. (8) : joint;} \end{cases}$$

9: **return** t_i for $\forall p_i \in \mathcal{P}$ in three strategies;

C. Joint Discount Allocation

By combining the individual-loss-based strategy and the individual-contribution-based strategy, we propose a new discount allocation strategy for passengers, called joint discount allocation strategy. To find the balance between the two strategies, we first introduce a parameter, denoted as $\lambda \in [0, 1]$. That is, the allocated discount of a passenger $p_i \in \mathcal{P}$ is proportional to the balanced value between its loss and contribution, i.e.,

$$t_i = \gamma \times T \times (\lambda \frac{\Delta l_i}{\sum_{p_k \in \mathcal{P}} \Delta l_k} + (1-\lambda) \frac{C_i}{\sum_{p_k \in \mathcal{P}} C_k}). \quad (8)$$

Since the individual-loss-based discount allocation strategy is only suitable for passengers, the joint strategy is also only applicable to passengers. By controlling the balanced parameter λ, the passengers with high contribution and large loss will receive large discount, and they will further participate in the privacy-preserving ride-hailing system, so that the system is long-term profitable.

D. The Detailed Algorithms

Based on the above strategies, we design the discount allocation algorithms for passengers and drivers, respectively, as shown in Algorithms 2 and 3. First, we introduce Algorithm 2. We initialize the discount values for all passengers and compute the social welfare value of the existing assignment results in Steps 1-2. Then, each passenger (e.g., p_i) calculates its individual loss and sends it to the platform in Steps 3-4. At the same time, the platform re-conducts the matching between drivers and passengers excluding p_i and then gets a new social welfare value in Step 5. Based on this, the contribution of this passenger p_i can be obtained in Step 6. Next, the allocated discount for each passenger p_i is determined in Step 8. Here, we present the allocated discount values based on the three

Algorithm 3 Discount Allocation Algorithm for Drivers

Input: $\mathcal{P}$, $\mathcal{D}$, γ, κ, T
Output: t_j for $\forall d_j \in \mathcal{D}$.
1: Initialization: $t_j = 0$ for $\forall d_j \in \mathcal{D}$;
2: Platform computes the total social welfare, i.e., W;
3: **for** $d_j \in \mathcal{D}$ **do**
4: Platform re-matches passengers $\mathcal{P}$ and drivers $\mathcal{D}/\{d_j\}$, and further computes the new social welfare, i.e., W_{-d_j};
5: Platform computes the contribution $G_j = W - W_{-d_j}$;
6: **for** $d_j \in \mathcal{D}$ **do**
7: Platform calculates the discount for d_j based on Eq.(7);
8: **return** t_j for $\forall d_j \in \mathcal{D}$;

proposed strategies. We finally output the discount results for passengers in Step 9.

Second, we introduce the discount allocation algorithm for drivers, i.e., Algorithm 3. Similar to Algorithm 2, in Steps 1-2, we first initialize the algorithm and meanwhile compute the social welfare based on the already assigned results. Next, for each driver $d_j \in \mathcal{D}$, we remove d_j and re-conduct the matching between drivers $\mathcal{D}/\{d_j\}$ and passengers $\mathcal{P}$ in Step 4. After calculating the contribution of each driver in Step 5, we obtain the corresponding discount according to Eq. (7), in Steps 6-7. At last, we output the results in Step 8.

E. Example

To better understand the discount allocation strategies, we use an example shown in Table II to illustrate the allocation procedure. Same as the former example in Section IV, we suppose 3 passengers and 3 drivers in the system. Here, the true and false locations of passengers are included, as shown in Fig. 4. Moreover, the distance values between drivers and passengers are shown in Fig. 5. Then, the discount allocation procedure is conducted as follows.

Note that the locations of drivers in the system are public to passengers. For a passenger (e.g., p_1) in the example, it can calculate the distance between him and the closest driver (i.e., d_1) which equals $2\sqrt{2}$. However, its actually assigned driver is d_3, and the corresponding distance equals $\sqrt{10}$. So the individual loss for p_1 is calculated by $\sqrt{10}-2\sqrt{2} \approx 0.33$. In the same way, we can compute the individual loss for other passengers, as shown in Table II. Note that here the total profits that will be allocated to all passengers are determined, i.e., γT. Thus, we calculate the allocated discount for each passenger according to Eq. (5), and display the results in Table II.

On the other hand, we also analyze the results in the individual-contribution-based allocation strategy. In the system, the social welfare is reversely proportional to the total distance, so we use the opposite of total distances to denote the social welfare. Note that the individual-contribution-based allocation strategy works according to the false locations of passengers. This means that the computation of an individual contribution is always based on Fig. 4(b). We first calculate the total social welfare based on all users, i.e., $W = (-2) + (-\sqrt{2}) + (-\sqrt{5}) \approx -5.65$. Then, for the passenger p_1, we re-conduct the optimal matching based on all passengers excluding p_1, and get the new assignment result $\{\langle p_2, d_1\rangle, \langle p_3, d_2\rangle\}$. Based on this, we re-calculate the social welfare, i.e., $W_{-p_1} = (-\sqrt{2}) + (-\sqrt{5}) \approx -3.65$. In this case, the contribution of p_1, defined as the difference between the social welfare values including p_1 and excluding p_1, can be determined as $C_{p_1} = |W - W_{-p_1}| = 2$. In the same way, the contribution of p_2 and p_3 is calculated, as shown in Table II. Next, according to Eq. (6), the discount allocated to each individual user is determined.

TABLE II
THE EXAMPLE OF DISCOUNT ALLOCATION STRATEGIES.

Passengers	Individual-loss-based discount allocation		Individual-contribution-based discount allocation	
	loss	discount	contribution	discount
p_1	0.33	$0.21\lambda T$	2	$0.35\lambda T$
p_2	1	$0.64\lambda T$	1.41	$0.25\lambda T$
p_3	0.24	$0.15\lambda T$	2.24	$0.4\lambda T$

Additionally, based on the above two strategies, it is easy to compute the discount in the joint allocation strategy. The specific discount value of each passenger depends on the balance parameter λ. Moreover, we can also calculate the discount values for all drivers based on Eq. (7), by using the individual-contribution-based allocation strategy. The procedure is the same as the discount allocation for passengers in the example, so we will omit the detailed calculation here.

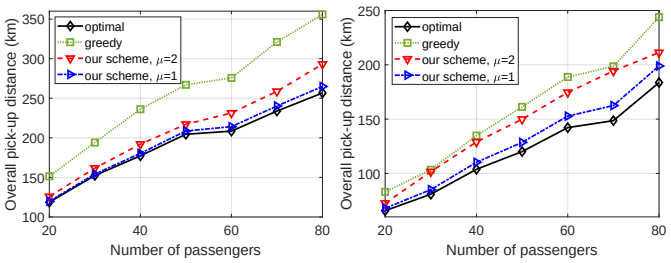
VI. EXPERIMENT

In this section, we evaluate the performances of the proposed algorithms. We conduct the simulations on a computer with Inter(R) Core(TM) i7-8700 CPU @3.20GHz and 32GB RAM under a Windows platform. Moreover, all simulations are implemented in Matlab.

A. Experiment setup

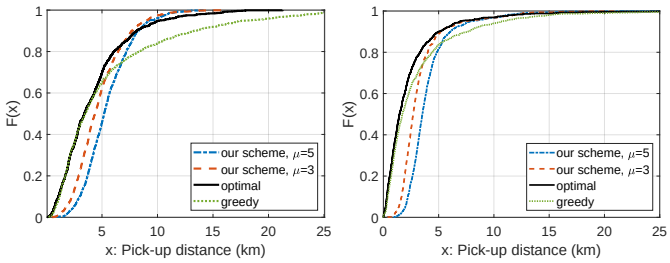
In the experiment, we use both synthetic and real-world datasets. In the synthetic dataset, the locations of passengers and drivers are randomly generated with uniform distribution. Specifically, we first generate a planar size in 30×30. Then, we generate driver and passenger locations in the area, where each location is represented by a 2D coordinate. The distributions of the coordinate values are uniform. In the real-world dataset, these locations are extracted from the order and trace data in Chengdu, China from Didi Inc. The dataset is available at [26]. To best of our knowledge, there is no available dataset that contains privacy requirements of passengers. Therefore, we assume the privacy requirements obey normal distribution. The mean of the distribution is denoted as μ, and the standard deviation is set as $\mu/3$, which could guarantee that 99.7% of the generated privacy requirement is positive by expectation. If a negative privacy requirement is generated, we manually adjust it to 0. In both datasets, we set the number of drivers to the same as that of passengers.

In the first group of experiments, we investigate the matching performance of our scheme on both datasets. We first compare our bipartite-matching based scheme with the simple spatial cloaking approach in which each passenger greedily chooses the nearest available driver. The comparison algorithm



(a) Results on the synthetic dataset (b) Results on the real-world dataset

Fig. 7. Comparison on the overall pick-up distance.



(a) Results on the synthetic dataset (b) Results on the real-world dataset

Fig. 8. Comparison on the pick-up distance distribution.

is denoted as greedy. We also compare our scheme with the optimal solution in which the actual locations of passengers are used in the matching.

Then, we simulate the discount allocation strategies. Note that the location information of drivers is public to the passengers, so the loss of each passenger is calculated easily. Then, the individual-contribution-based allocations for drivers and passengers are computed according to two matching results. In the joint discount allocation algorithm, we set the parameter λ from $\{0.3, 0.5, 0.7\}$, and the default λ is set as 0.5. In addition, the total discounts allocated to drivers or passengers are proportional to the number of drivers or passengers.

B. Simulation Results

Fig. 7 shows the comparison of different algorithms on the overall pick-up distances. Fig. 7(a) illustrates simulation results on the synthetic dataset. From the figure, we can find out that our ride matching algorithm outperforms the simple spatial cloaking approach (denoted as greedy). The reason is that each driver could only be chosen once, and the global matching based algorithms could coordinate between passengers and minimize the overall pick-up distance. If actual locations of passengers are known, the bipartite matching algorithm should achieve the optimal value as the black solid line shown in the figure. When the passenger locations are protected by cloaking regions, the matching performance decreases as the red and blue lines shown in the figure. By comparing the red line and blue line, we can verify that larger privacy requirements would result in larger overall pick-up distance. Fig. 7(b) illustrates simulation results on the real-world dataset. It shares similar trends with the results on the synthetic dataset. The difference is that the effect of privacy requirements is more obvious. When changing the value of μ from 1 to 2, the relative difference between the red line and the blue line is larger in Fig. 7(b) than that in Fig. 7(a). Although the effect of the value of μ is more obvious, our scheme still outperforms the simple spatial cloaking approach.

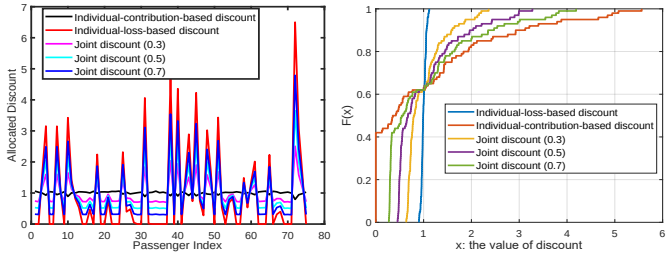
Fig. 8 shows the comparison of different algorithms on the pick-up distance distribution. Fig. 8(a) and (b) plot the cumulative distribution function of the pick-up distances. From Fig. 8(a), we can find out that 60% of the pick-up distances are less than 4.4km when the simple spatial cloaking approach is used. The corresponding value of our scheme is 4.9km when $\mu = 3$, which is larger. In contrast, when investigating the 80% of the pick-up distances, they are less than 8.4km when the simple spatial cloaking approach is used, and the corresponding value of our scheme is 6.4km when $\mu = 3$. This shows that letting passengers choose the neatest driver could benefit some passengers whose pick-up distance is relatively small, while it also makes negative effects on the pick-up distances of some other passengers. When using our scheme, the pick-up distances are more concentrated compared with the simple spatial cloaking approach. We can find out the similar conclusion from Fig. 8(b).

We evaluate the allocated discount for each passenger based on three allocation algorithms, as shown in Fig. 9 (a). We see that the allocated discount differences among all passengers in the individual-loss-based algorithm are larger than those of other algorithms, and the discount differences in the individual-contribution-based algorithm are smallest. This means that the loss of each passenger dominates in the joint discount allocation algorithm. Moreover, we present the cumulative probability distribution of the allocated discount for three algorithms in Fig. 9 (b). We also show the discount results when changing the parameter λ. We find that in the individual-loss-based allocation algorithm, the largest discount value for one passenger is about 5.5 and there are about 40% passengers who get no discount. While in the individual-contribution-based algorithm, the differences in the allocated discounts for passengers are small. When we change the parameter λ, all passengers will get certain discounts.

Also, we show the variance of the additional expense in Definition 1 in Fig. 10 (a). We find that the individual-loss-based algorithm always achieves the minimum variance value while the individual-contribution-based algorithm gets the maximum variance value. When we increase the number of passengers and drivers, the variance values of all algorithms will decrease. This is because the global matching result will get better when more passengers and drivers join. These simulations are consistent with our theoretical analysis. On the other hand, we also evaluate the discount allocation for each driver in Fig. 10 (b). Since the true locations of passengers are not invisible to drivers, only the individual-contribution-based algorithm can be applied for drivers. We thus find that the allocated discount values for each passenger change a little.

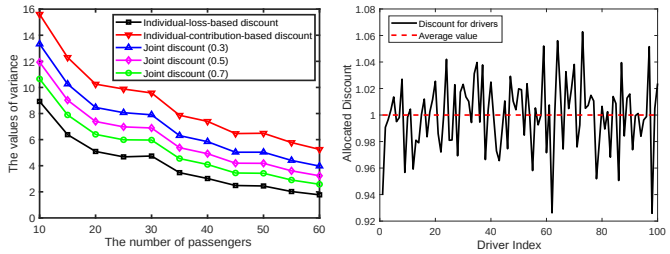
VII. Conclusion

In this paper, we introduce a privacy-preserving order dispatch system for ride-hailing services. Different from the previous approaches that let passengers choose the desired drivers, we investigate the approach that lets the service



(a) Number of passengers: 75 (b)Cumulative probability distribution

Fig. 9. The discount distribution for passengers.



(a) Simulation of fairness (b) The discount for drivers

Fig. 10. The discount distribution for passengers and drivers.

provider match passengers and drivers in a centralized way. Our approach can not only prevent the location inference attacks introduced in [11], but also avoid the large communication overhead introduced by encryption. Based on our approach, we propose to maximize the social welfare (or minimize the overall pick-up distance) of the matching. Under the privacy requirement S_i, we show that the overall pick-up distance of our matching result is upper bounded by $OPT + \sqrt{2}\sum_{p_i \in \mathcal{P}} \sqrt{S_i}$, while the strong privacy defined in [11] is kept. In addition, we introduce three discount allocation schemes to make up for the loss of all individuals caused by global matching. Experiments on both synthetic and real-world datasets show the efficiency of our scheme.

ACKNOWLEDGEMENT

This research was supported in part by NSF grants CNS 1824440, CNS 1828363, CNS 1757533, CNS 1618398, CNS 1651947, and CNS 1564128.

REFERENCES

[1] Y. Duan, T. Mosharraf, J. Wu, and H. Zheng, "Optimizing carpool scheduling algorithm through partition merging," in *IEEE ICC*, 2018.

[2] G. Gao, M. Xiao, and Z. Zhao, "Optimal multi-taxi dispatch for mobile taxi-hailing systems," in *IEEE ICPP*, 2016, pp. 294–303.

[3] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *2011 IEEE symposium on security and privacy*. IEEE, 2011, pp. 247–262.

[4] G. Meyer-Lee, J. Shang, and J. Wu, "Location-leaking through network traffic in mobile augmented reality applications," in *IEEE IPCCC*, 2018.

[5] E. Weise and J. Guynn, "Uber tracking raises privacy concerns," https://www.usatoday.com/story/tech/2014/11/19/uber-privacy-tracking/19285481/.

[6] M. L. Damiani, E. Bertino, C. Silvestri *et al.*, "The probe framework for the personalized cloaking of private locations." *Trans. Data Privacy*, vol. 3, no. 2, pp. 123–148, 2010.

[7] M. Xue, P. Kalnis, and H. K. Pung, "Location diversity: Enhanced privacy protection in location based services," in *International Symposium on Location-and Context-Awareness*. Springer, 2009, pp. 70–87.

[8] A. Pham, I. Dacosta, G. Endignoux, J. R. T. Pastoriza, K. Huguenin, and J.-P. Hubaux, "Oride: A privacy-preserving yet accountable ride-hailing service," in {*USENIX*} *Security*, 2017, pp. 1235–1252.

[9] U. M. Aïvodji, K. Huguenin, M.-J. Huguet, and M.-O. Killijian, "Sride: A privacy-preserving ridesharing system," in *ACM WiSec*, 2018, pp. 40–50.

[10] Y. He, J. Ni, X. Wang, B. Niu, F. Li, and X. Shen, "Privacy-preserving partner selection for ride-sharing services," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 5994–6005, 2018.

[11] Y. Khazbak, J. Fan, S. Zhu, and G. Cao, "Preserving location privacy in ride-hailing service," in *IEEE CNS*, 2018.

[12] F. Aurenhammer, "Voronoi diagramsa survey of a fundamental geometric data structure," *ACM Computing Surveys (CSUR)*, vol. 23, no. 3, pp. 345–405, 1991.

[13] D. Sánchez, S. Martínez, and J. Domingo-Ferrer, "Co-utile p2p ridesharing via decentralization and reputation management," *Transportation Research Part C: Emerging Technologies*, vol. 73, pp. 147–166, 2016.

[14] P. Goel, L. Kulik, and K. Ramamohanarao, "Optimal pick up point selection for effective ride sharing," *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 154–168, 2017.

[15] C. Dai, X. Yuan, and C. Wang, "Privacy-preserving ridesharing recommendation in geosocial networks," in *CSoNet*. Springer, 2016, pp. 193–205.

[16] U. M. Aïvodji, S. Gambs, M.-J. Huguet, and M.-O. Killijian, "Meeting points in ridesharing: A privacy-preserving approach," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 239–253, 2016.

[17] H. Li, H. Zhu, S. Du, X. Liang, and X. S. Shen, "Privacy leakage of location sharing in mobile social networks: Attacks and defense," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 646–660, 2018.

[18] N. Zhang, S. Zhong, and L. Tian, "Using blockchain to protect personal privacy in the scenario of online taxi-hailing." *International Journal of Computers, Communications & Control*, vol. 12, no. 6, 2017.

[19] A. Pham, I. Dacosta, B. Jacot-Guillarmod, K. Huguenin, T. Hajar, F. Tramèr, V. Gligor, and J.-P. Hubaux, "Privateride: A privacy-enhanced ride-hailing service," *Proceedings on Privacy Enhancing Technologies*, vol. 2017, no. 2, pp. 38–56, 2017.

[20] X. Liang, X. Li, R. Lu, X. Lin, and X. Shen, "Udp: Usage-based dynamic pricing with privacy preservation for smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 141–150, 2013.

[21] X. Zhuo, W. Gao, G. Cao, and Y. Dai, "Win-coupon: An incentive framework for 3g traffic offloading," in *IEEE ICNP*, 2011, pp. 206–215.

[22] G. Gao, M. Xiao, J. Wu, L. Huang, and C. Hu, "Truthful incentive mechanism for nondeterministic crowdsensing with vehicles," *IEEE Transactions on Mobile Computing*, vol. 17, no. 12, pp. 2982–2997, 2018.

[23] B. Edelman, M. Ostrovsky, and M. Schwarz, "Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords," *American economic review*, vol. 97, no. 1, pp. 242–259, 2007.

[24] W. Vickrey, "Counterspeculation, auctions, and competitive sealed tenders," *The Journal of finance*, vol. 16, no. 1, pp. 8–37, 1961.

[25] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar, "Preserving user location privacy in mobile data management infrastructures," in *International Workshop on Privacy Enhancing Technologies*. Springer, 2006, pp. 393–412.

[26] Data source: Didi chuxing gaia open dataset initiative. [Online]. Available: https://gaia.didichuxing.com