Duan Y, Gao GJ, Xiao MJ *et al.* Cloaking region based passenger privacy protection in ride-hailing systems. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 35(3): 629–646 May 2020. DOI 10.1007/s11390-020-0256-1

# Cloaking Region Based Passenger Privacy Protection in Ride-Hailing Systems

Yubin Duan[1], *Student Member, IEEE*, Guo-Ju Gao[2,*], *Student Member, IEEE*, Ming-Jun Xiao[2], *Member, IEEE* and Jie Wu[1], *Fellow, IEEE*

[1] *Department of Computer and Information Sciences, Temple University, Pennsylvania 19122, U.S.A.*

[2] *School of Computer Science and Technology, University of Science and Technology of China, Hefei 230036, China*

E-mail: yubin.duan@temple.edu; gaoguoju@mail.ustc.edu.cn; xiaomj@ustc.edu.cn; jiewu@temple.edu

Received December 30, 2019; revised March 11, 2020.

**Abstract** With the quick development of the sharing economy, ride-hailing services have been increasingly popular worldwide. Although the service provides convenience for users, one concern from the public is whether the location privacy of passengers would be protected. Service providers (SPs) such as Didi and Uber need to acquire passenger and driver locations before they could successfully dispatch passenger orders. To protect passengers' privacy based on their requirements, we propose a cloaking region based order dispatch scheme. In our scheme, a passenger sends the SP a cloaking region in which his/her actual location is not distinguishable. The trade-off of the enhanced privacy is the loss of social welfare, i.e., the increase in the overall pick-up distance. To optimize our scheme, we propose to maximize the social welfare under passengers' privacy requirements. We investigate a bipartite matching based approach. A theoretical bound on the matching performance under specific privacy requirements is shown. Besides passengers' privacy, we allow drivers to set up their maximum pick-up distance in our extended scheme. The extended scheme could be applied when the number of drivers exceeds the number of passengers. Nevertheless, the global matching based scheme does not consider the interest of each individual passenger. The passengers with low privacy requirements may be matched with drivers far from them. To this end, a pricing scheme including three strategies is proposed to make up for the individual loss by allocating discounts on their riding fares. Extensive experiments on both real-world and synthetic datasets show the efficiency of our scheme.

**Keywords** order dispatch, pricing, privacy, ride-hailing

## 1 Introduction

Nowadays, ride-hailing services have rapidly developed [1, 2]. Millions of users are attracted by the travel convenience provided by such a service. However, the public concerns about location privacy also rise along with the development of the ride-hailing service. Specifically, the service providers (SPs), such as Didi or Uber, could access and gather millions of travel traces per day. With novel data mining techniques, SPs have the ability to infer the private information of passengers by digging in the travel trace records. [3, 4] have shown that the SP could infer the living or working addresses of passengers, or collect their habits and interest. Passengers might suffer from location-based scams if such information is exposed to unauthorized organizations. In extreme cases, it might lead to economic or social reputation damage to passengers. Thus, it is necessary to protect the location privacy of passengers when providing ride-hailing services.

The existing privacy protection schemes for ride-hailing can generally be divided into the spatial cloaking based approaches [5, 6] and the homomorphic encryption based approaches [7–9]. In the spatial cloaking approach, cloaking regions rather than actual locations of

Regular Paper

Recommended by IEEE MASS 2019

A preliminary version was published in the Proceedings of IEEE MASS 2019.

This research was supported in part by the National Science Foundation of USA under Grant Nos. CNS 1824440, CNS 1828363, CNS 1757533, CNS 1618398, CNS 1651947, and CNS 1564128, the National Natural Science Foundation of China under Grant Nos. 61872330, 61572457, 61379132, and the National Natural Science Foundation of Jiangsu Province of China under Grant Nos. BK20191194 and BK20131174.

*Corresponding Author

passengers are reported to the SP. The passenger's actual location is indistinguishable in the corresponding cloaking region. To match passengers with drivers without knowing actual passenger locations, one existing solution is to let each passenger choose the nearest available driver. In this way, the SP could not know the actual locations of passengers. However, social welfare is not considered in this approach. Specifically, the overall pick-up distance is not optimized, and the efficiency of the ride-hailing system is not fully achieved. In the homomorphic encryption based approach, some encrypted or non-sensitive information could be attained by the SP, and the SP performs the calculation on these messages. For example, the SP might only know the distance between passengers and drivers without knowing their locations. It is hard for the SP to infer the actual passenger locations. The weakness of the homomorphic encryption based approach is that it usually leads to additional communication overheads. We aim to design a scheme that could overcome the shortfall of both approaches.

The existing cloaking region based approach also might be attacked by the model proposed in [10]. In the attacking model, the SP is assumed to be honest but curious, i.e., the SP would send accurate driver locations to passengers but try to infer passenger locations. The SP could use the Voronoi diagram [11] to infer passenger locations, since the SP knows that each passenger would choose the nearest available driver. Although [10] proposes an enhanced scheme to prevent such attacks, the authors of [10] did not consider the social welfare.

We aim to design a dispatch scheme which could maximize the social welfare (or minimize the overall pick-up distance) while ensuring privacy requirements of passengers and maintaining a low communication overhead. In this paper, a privacy-preserving order dispatch scheme based on spatial cloaking is introduced.

Different from the existing approaches, 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 [10]. The trade-off is that the performance of matching is affected, since the SP only knows the cloaking regions rather than the actual locations of passengers. We investigate the trade-off and prove that our scheme could achieve a theoretical bound on the social welfare under given privacy requirements.

On the other hand, maximizing the social welfare (i.e., minimizing the overall pick-up distance) may sacrifice some individuals' interest. The SP needs to allocate incentives (such as discounts on ride fees) to make up for their losses. For example, passengers with low privacy requirements may be matched with drivers who are far from them, while passengers with larger cloaking regions may be matched with relatively nearer drivers. In such cases, the low-privacy passengers could be charged more money than those high-privacy passengers, which is not reasonable. Therefore, we propose to make up for 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 for individual losses. The individual loss of each passenger is compared with the distance to the nearest driver. Allocating discounts would not leak the location privacy of passengers since the SP cannot determine which driver is the nearest to the passenger. It is challenging to determine a closed-form equation to describe the relationship between privacy requirements and their side effects on the 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 that 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 the locations of cloaking region centers and drivers. By using this scheme, the SP could not infer the 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 interest, 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 their ride.

This paper is an extended version of the conference paper [12] published in IEEE MASS 2019. Besides pas-

senger privacy, we consider the pick-up distances of drivers in an extended scheme. Specifically, the original matching scheme might lead to extremely long pick-up distances for drivers, which is not reasonable. In the extended scheme, we allow drivers to set up the maximum pick-up distance to avoid such cases. The matching process of the original scheme is modified accordingly. The trade-off is that it might result in unserved passengers. The extended scheme could be deployed when the number of drivers is more than that of passengers.

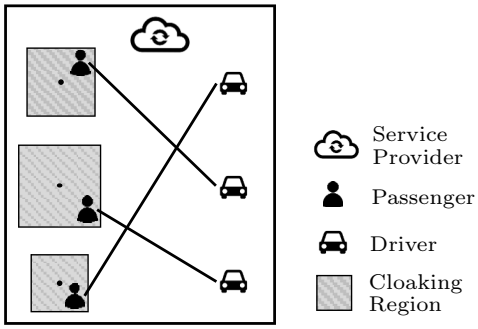


Fig.1. Application scenario.

Our contributions are summarized as follows.

• We propose a cloaking region based order dispatch scheme. It could prevent the attack introduced in [10], with no significant communication overhead.

• We further extend our scheme in consideration of drivers' pick-up distance. In the extended scheme, drivers are allowed to set limitations on pick-up distances.

• We evaluate the loss of social welfare caused by privacy requirements (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.

• Extensive simulations based on real-world and synthetic datasets are conducted to verify the significant performance of our algorithms.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 presents our system model. Section 4 shows the ride matching algorithm and analyzes its performance and privacy-preserving properties. Section 5 investigates an extended case where the drivers' pick-up distance constraints are considered. Section 6 introduces our discount allocation strategies. Section 7 simulates our approaches on both synthetic and real-world datasets. Finally, Section 8 concludes the paper.

## 2 Related Work

This section reviews the existing strategies of protecting location privacy in ride-hailing systems and approaches used for discount allocation.

### 2.1 Location Privacy in Ride-Hailing System

We first focus on researches about protecting the location privacy of each passenger in the ride-hailing system. The importance of this topic is increasingly driven by the public's rising attention to location privacy and the quick development of intelligent transportation systems [13]. Researchers proposed several general privacy-preserving approaches for pervasive computing, such as [14–16]. Rather than the general approaches, privacy protection schemes for ride-hailing systems are closer to our work. Therefore, we would mainly review these schemes.

We could divide the common privacy protection schemes for ride-hailing systems, such as [10, 17–22], into two groups. One group uses the location cloaking approach [5, 6, 10] to protect location privacy; while the other group is based on the homomorphic encryption approach [7–9]. 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 [23]. The SP would send locations of all available drivers in cloaking regions to passengers. Then, passengers can choose drivers based on some metrics. [23] proposes to let each passenger choose the nearest driver. However, [10] points out that the SP could infer actual passenger locations to a certain degree by using the Voronoi diagram [11]. To enhance the privacy level, [10] proposes to choose relatively nearer drivers with a 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 a 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.

### 2.2 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 [24] designed a usage-based dynamic pricing

scheme with privacy preservation for smart grids, in which they enabled the electricity price to correspond to the electricity usage in real time. Zhuo *et al.* in [25] studied the trade-off between the amount of traffic being offloaded and the users' satisfaction in 3G network, and further proposed 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 [26], such as generalized second-price auction [27], Vickrey-Clarke-Groves (VCG) auction [28]. 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 an amount equal to 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 the VCG auction while taking the fairness of discount into consideration.

## 3 Model

In this section, we first briefly describe the overview of the ride-hailing system. Then, notations used to model the system are introduced. The formulation of our problem is given at last.

### 3.1 Overview of Ride-Hailing Systems

The ride-hailing system consists of three parties: passengers, drivers, and the service provider (SP). Passengers have travel demands and would contact the SP to request drivers. In this paper, each passenger also has their own privacy requirements. We assume they are willing to afford additional costs caused by their privacy requirements. Drivers would pick up passengers from the origins and send them to the destinations. We assume that idle drivers would share their locations with the SP without privacy requirements. Instead, drivers could set limitations on pick-up distances, i.e., the longest distance they could afford to pick up the passengers assigned by the SP. The SP would gather passenger requests and the locations of idle drivers, and match passengers and drivers.

The SP is the potential attacker of the ride-hailing system. We assume the SP is honest-but-curious. Specifically, the SP would follow the order dispatch scheme and would not integrate malicious plugins in its mobile apps for either the passenger or the rider side, since the applications released by the SP usually would be reviewed by app stores. The SP is just curious about passengers' origins and destinations from which the SP could infer valuable information such as the hobbies of the passengers. In addition, we assume drivers would not cooperate with the SP. Although drivers know the actual origins and destinations of passengers, they have no other information about passengers such as their payment information.

### 3.2 Existing Attack Model

An attack model on passengers' location privacy in the ride-hailing system is proposed in [10]. The attack model assumes each passenger has a cloaking region in which his/her location is indistinguishable with other locations in the region. The service provider (SP) would send each passenger the locations of drivers in the passenger's clocking region. Each passenger would choose the nearest driver. It seems that the SP could not directly access passengers' locations except their cloaking regions. However, toward this dispatch scheme, the SP could improve the inference of passengers' locations by using the Voronoi diagram. Specifically, the SP could construct a Voronoi diagram based on all drivers' locations to launch the attack. In the Voronoi diagram, each driver's location would be enclosed by a Voronoi polygon which contains all the locations whose distance to the driver is less than or equal to its distance to any other driver. Based on each passenger's selection, the SP could infer that the passenger's location is within the chosen driver's Voronoi polygon. If the Voronoi polygon is smaller than the passenger's cloaking region, then the SP could reduce the passenger's cloaking region into its intersection of the Voronoi polygon.

### 3.3 Notations

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

To protect the location privacy, each passenger would construct a cloaking region based on their pri-

vacy requirements. The cloaking region [29] 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 circled 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 choose 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.

**Table 1**. Description of Commonly Used Notations

| Variable | Description |
|---|---|
| $\mathcal{P}$, $\mathcal{D}$ | Sets of passengers and drivers, respectively |
| $i$, $j$ | Indexes of passengers and drivers, respectively |
| $R_i$ | $p_i$'s privacy requirement |
| $S_i$ | $p_i$'s cloaking region |
| $p_i'$ | $p_i$'s reported location |
| $\mathcal{P}'$ | Set of reported locations |
| $dis(\cdot,\cdot)$ | Distance function |
| $\eta_j$ | Distance limitation of driver $d_j$ |
| $T$ | Total additional fares for privacy |
| $\gamma$, $\kappa$ | Parameters for the discount allocation |
| $\Delta l_i$ | Local distance loss of passenger $p_i$ |
| $W$ | Total other passengers' social welfare (SW) |
| $W_{-p_i}$ | Total other SW on matching excluding $p_i$ |
| $C_i$ | Global SW loss of passenger $p_i$ |
| $\lambda$ | Parameter to balance LD and SW loss |
| $G_j$ | Total other drivers' SW after excluding $d_j$ |

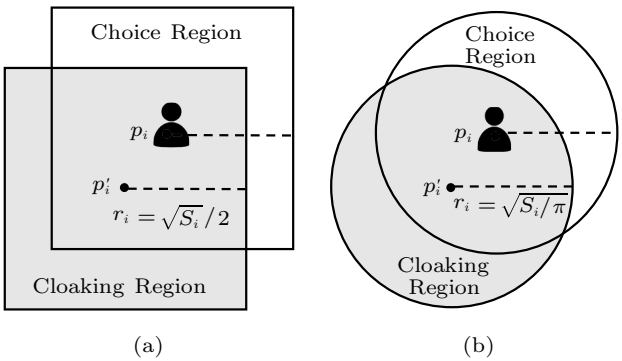


Fig.2. Illustration of the cloaking region [12]. (a) Square. (b) Circle.

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 6. The performance of the ride matching scheme is evaluated by the social welfare (or the overall pick-up distance). Formally, let $dis(\cdot,\cdot) : \mathbb{R}^2 \mapsto \mathbb{R}$ denote the distance function. The pick-up 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 pick-up distance. It is because a longer pick-up distance corresponds to lower social welfare.

### 3.4 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 follows:

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

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

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

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

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

## 4 Privacy Preserving Dispatch Process

This section introduces our cloaking region based privacy protection scheme. We first sketch our system framework, and then describe the details of our order matching process. Finally, we analyze the performance

of the matching process and the privacy-preserving property of our scheme.

### 4.1 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. Also, drivers could set maximum pick-up distances based on their preference, which is described in Section 5. In this subsection, we consider the simple version that ignores limitations on pick-up distances. The privacy requirements of passengers are represented by the sizes of cloaking regions.

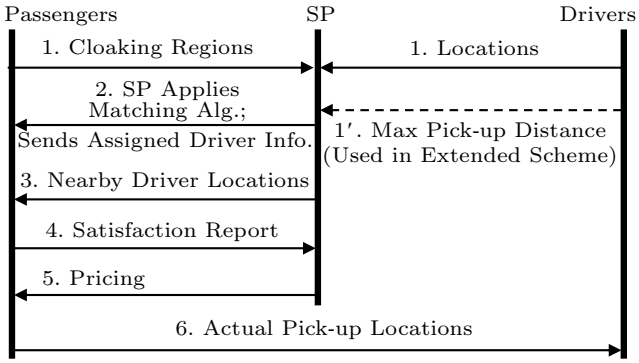


Fig.3. System framework.

Then, in the second step, the SP would match passengers with drivers by using the weighted bipartite matching algorithm [2], and send the matching results to passengers. Details of the matching process are introduced in Subsection 4.2. 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 locations of drivers around the cloaking regions to passengers. 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 (introduced in Section 5) 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. The procedure of our order dispatch scheme is shown in Algorithm 1.

In the scheme, only idle drivers would share their locations with the SP. In addition, once a driver is assigned to pick up a passenger, he/she stops sharing his/her location. After arriving at the passenger's destination, the driver would start sharing his/her location after a random-length time period. These could prevent the SP from inferring private passenger locations from the traces of idle drivers.

**Algorithm 1.** Order Dispatch Scheme

**Input:** passengers' reported locations $\mathcal{P}'$ and privacy requirements $\{S_i|p'_i \in \mathcal{P}'\}$, 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 passenger along with the locations of drivers near the cloaking region;
4: Receive satisfactions from passengers;
5: Pricing for passengers $\leftarrow$ Discount Allocation Algorithm for Passengers$(\mathcal{P}', \mathcal{D}, \gamma, \kappa, \lambda, T)$ (Algorithm 3); pricing for drivers $\leftarrow$ Discount Allocation Algorithm for Drivers$(\mathcal{P}', \mathcal{D}, \gamma, \kappa, \lambda, T)$ (Algorithm 4);
6: **return** $M$;

### 4.2 Matching with Cloaking Regions

In our scheme, to protect the location privacy of passengers in the ride-hailing systems, we propose to let passengers upload the cloaking regions instead of their precise locations to the SP. In our matching process, we propose using the centers of the cloaking regions as obfuscated locations of passengers. Then, we apply the weighted bipartite matching algorithm with the obfuscated locations of passengers and driver locations. Although the matching result is not optimal with respect to actual passenger 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 $\mathcal{D}$. The weight of the edge between $p' \in \mathcal{P}'$ and $d \in \mathcal{D}$ is the negation of the geographical distance between the obfuscated location $p'$ and the 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 the sum of distances between passengers and drivers.

An example of our dispatch scheme is shown in Figs.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 actual passenger

locations and drivers. The distance between the locations used in the example is given in Fig.5.

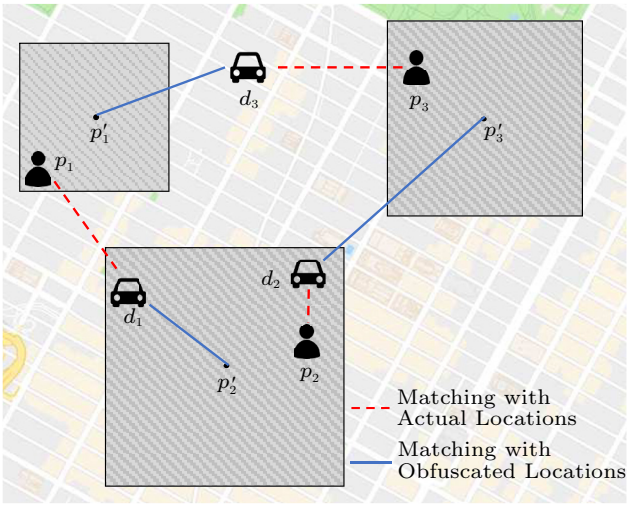


Fig.4. Effect of cloaking regions in the matching process [12].

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

(a)

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

(b)

Fig.5. Distance table of the example [12]. (a) Distances between actual locations and drivers. (b) Distances between reported locations and drivers.

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 $dis(p_1, d_3) + dis(p_2, d_1) + dis(p_3, d_2) = \sqrt{10}+2+\sqrt{5} = 7.40$. Note that we should use the actual locations rather than the 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, not the 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 Subsection 4.3.

### 4.3 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 actual passenger locations in the bipartite matching. In our scheme, the SP could only perform the bipartite matching based on the obfuscated locations of passengers. Let $M$ denote the overall distance between the obfuscated locations of passengers and the 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 distances between the actual location of passenger $p_i$ and the location of driver $d'_j$. Formally, $M' = \sum_{p_i \in \mathcal{P}} dis(p_i, d'_j)$. Fig.6 illustrates the meaning of these notations, where $OPT$ means the matching with actual locations, $M$ represents the matching with obfuscated locations, and $M'$ shows the pick-up distance (picking up at actual passenger locations).

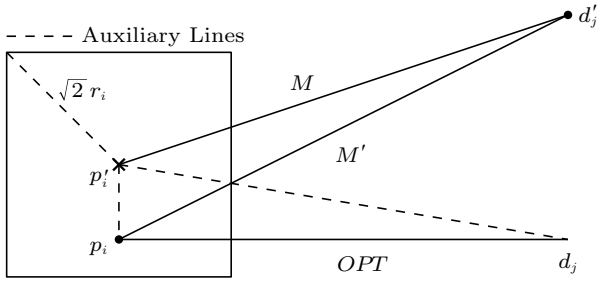


Fig.6. Illustration of the proof [12].

**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 $dis(p_i, p'_i) \leqslant \sqrt{2}r_i$.

Based on the triangle inequality, we have that

$$\begin{aligned} M' &= \sum_{p_i \in \mathcal{P}} dis(p_i, d'_j) \\ &\leqslant \sum_{p_i \in \mathcal{P}} (dis(p'_i, d'_j) + dis(p_i, p'_i)) \\ &= M + \sum_{p_i \in \mathcal{P}} dis(p_i, p'_i). \end{aligned}$$

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

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

Above all, we have

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

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* [10].

*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 [10], we could conclude that the strong privacy holds. □

## 5 Dispatch with Pick-up Distance Constraints

In this section, we take the efficiency of the taxi-hailing system into consideration. In reality, each driver is not willing to pick up a passenger from a long distance, while the passengers are reluctant to be allocated to the drivers at a long distance. In such a case, the distance between a driver and the allocated passenger in our privacy-preserving taxi-hailing system should not be greater than a threshold.

### 5.1 Distance Constraint

In our scheme, we globally match passengers and drivers. It is possible that a driver needs to pick up a passenger who is far from him/her. In our preliminary experiment, we notice that some actual pick-up distances are extremely large, which is not reasonable for drivers. To prevent such a case, we add the distance constraint to our problem. Specifically, let drivers upload the maximum pick-up distances they could accept when they join the system. Let $\eta_j$ denote the distance limitation set by driver $d_j$. If passenger $p_i$ is assigned to driver $d_j$, then the pick-up distance $dis(p_i, d_j)$ should not exceed the distance limitation $\eta_j$. Therefore, the distance constraint can be formulated as

$$dis(p_i, d_j) \leqslant \eta_j, \forall d_j \in \mathcal{D}. \tag{5}$$

### 5.2 Extended Scheme

To adjust our scheme with the additional constraint, we improve the bipartite matching algorithm used in our scheme. The insight is to remove edges which break the distaste constraint in the bipartite graph $G$. Note that the passenger's actual location is unknown to the SPs. It is not trivial to determine whether the actual pick-up distance exceeds the distance constraint or not. Recall that in the bipartite graph $G$, $E = -dis(\mathcal{P}' \times \mathcal{D})$, i.e., we use passengers' reported locations $\mathcal{P}'$ to determine edge weights. The weights reveal the distance between drivers and "fake" locations of passengers. The actual pick-up distances should be $dis(\mathcal{P} \times \mathcal{D})$. Therefore, we could not directly delete edges such that $dis(p'_i, d_j) > \eta_j, \forall p'_i \in \mathcal{P}', \forall d_j \in \mathcal{D}$. We need to infer the actual pick-up distance without leaking passengers' location privacy.

Thanks to the geometry properties of the Euclidean space, we could get the upper bound of the pick-up distance when sizes of cloaking regions are known. Since the privacy requirement $S_i$ of each passenger $p_i$ is known to the SP, the range of actual pick-up distances could be calculated based on the triangle inequality. For consistency, we assume the shape of cloaking regions is square in the following analysis, which could also be easily extended to the circle cloaking regions. Specifically, based on the privacy constraint in (3), we know that $dis(p'_i, p_i) \leqslant \sqrt{2}r_i = \sqrt{S_i/2}$. According to the triangle inequality, we have that $dis(p_i, d_j) \leqslant dis(p'_i, d_j) + dis(p'_i, p_i)$ or $dis(p_i, d_j) \leqslant dis(p'_i, d_j) + \sqrt{S_i/2}$, i.e., the actual pick-up distance is upper bounded. It provides a bridge to transfer the distance constraint on the actual pick-up distance $dis(p_i, d_j)$ to the constraint on edge weights in the bipartite graph $G$. If we have that $dis(p'_i, d_j) \leqslant \eta_j - \sqrt{S_i/2}$, then we could guarantee that $dis(p_i, d_j) \leqslant \eta_j$. To show it, if $dis(p'_i, d_j) \leqslant \eta_j - \sqrt{S_i/2}$, then we have that $dis(p'_i, d_j) + \sqrt{S_i/2} \leqslant \eta_j$. Also, we have

shown that $dis(p_i, d_j) \leqslant dis(p'_i, d_j) + \sqrt{S_i/2}$. Therefore, $dis(p_i, d_j) \leqslant \eta_j$.

To satisfy the distance constraint, we modify the edge set $E$ of the bipartite graph $G$ used for matching. Edges that could potentially lead to the violation of the distance constraint are deleted. Let $E'$ denote the edge set after removing these edges. Formally, $E' = E \setminus \{(p'_i, d_j) | dis(p'_i, d_j) > \eta_j - \sqrt{S_i/2}\}$ or $E' = \{(p'_i, d_j) | dis(p'_i, d_j) \leqslant \eta_j - \sqrt{S_i/2}\}$. Then, the modified edge set $E'$ could form a new graph $G' = (V, E')$. Based on the previous analysis, performing the bipartite matching on $G'$ could guarantee the distance constraint of drivers. The modified scheme is summarized as Algorithm 2.

**Algorithm 2.** Order Dispatch with the Distance Constraint

**Input:** passengers' reported locations $\mathcal{P}'$ and privacy requirements $\{S_i | p'_i \in \mathcal{P}'\}$, driver locations $\mathcal{D}$ and distance limitations $\{\eta_j | d_j \in \mathcal{D}\}$

**Output:** dispatch results with the distance constraint

1: Construct a weighted bipartite matching graph $G = (V, E)$. $V = \mathcal{P}' \bigcup \mathcal{D}$, $E = -dis(\mathcal{P}' \times \mathcal{D})$;
2: Update the edge set $E$ based on distance limitations $\{\eta_j | d_j \in \mathcal{D}\}$. $E' \leftarrow E \setminus \{(p'_i, d_j) | dis(p'_i, d_j) > \eta_j - \sqrt{S_i/2}\}$;
3: $M \leftarrow$ weighted bipartite matching on $G' = (V, E')$;
4: Same as steps 3–5 in Algorithm 1;
5: **return** $M$;

Allowing drivers to set up the maximum pick-up distances could avoid unreasonable long-distance pick-ups for drivers, while the trade-off is a lower ratio of passengers that could be picked up. For example, when drivers' pick-up distances are small, a passenger who is far from these available drivers would be assigned with no driver. The passenger has to wait for the next round of order dispatch. From this point of view, the extended scheme is suitable to use in the case that the number of drivers is more than the number of passengers.

## 6 Discount Allocation Strategies

As we have mentioned before, the proposed privacy-preserving order-dispatch system focuses 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/her, and each driver wants to pick up the closest passenger. Although the proposed algorithm can obtain a fine global performance, the individual users including passengers and drivers may suffer some loss. Therefore, in order to stimulate individual users to participate in the privacy-preserving order-dispatch system, we should give each participant 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 2. Here, the specific values of $\gamma$ and $\kappa$ are determined after three-party (i.e., system, passengers, and drivers) negotiation.

**Table 2**. Profits Share

| Party | Amount |
|---|---|
| Total profits | $T$ |
| Passengers | $\gamma \times T$ |
| Drivers | $(1-\gamma)\kappa \times T$ |
| SP | $(1-\gamma)(1-\kappa) \times T$ |

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, the local distance (LD) loss based discount allocation strategy, the global social welfare (SW) loss based discount allocation strategy, and the joint discount allocation strategy. Here, since the exact locations of passengers are unknown to drivers, the LD loss based and joint discount allocation strategies cannot be applied for drivers.

### 6.1 Local Distance Loss Based Discount Allocation

We first design a discount allocation strategy from the perspective of individual distance loss. Compared with 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 (called local distance loss). 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 a local distance (LD) 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 locations of $p_i$ and $d'_j$ can be calculated. Thus, it is easy to compute the difference (i.e., the LD loss) between the actual distance and the nearest distance for $p_i$, denoted as $\Delta l_i$. Note that $\Delta l_i$ here is a non-negative value, i.e., $\Delta l_i \geqslant 0$. 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, larger $\Delta l_i$ indicates more loss for $p_i$. Also, the system must compensate more money for the passengers with larger loss. Based on this observation, we design the LD loss based discount allocation strategy as follows.

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

$$t_i = \gamma \times T \times \frac{\Delta l_i}{\sum_{p_k \in \mathcal{P}'} \Delta l_k}. \tag{6}$$

This intuition-based allocation strategy can ensure 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 $\Delta_i$. 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 $\Delta_i = \Delta l_i - \rho \times t_i$ where $\rho$ denotes the balanced parameter.

**Definition 1**. *Let $\sigma$ 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 $\sigma$ indicates that the additional expense of passengers has little difference. Thus, the smaller $\sigma$, the fairer this system.*

In the LD loss based discount allocation strategy, the passengers with larger loss will receive more discount. Obviously, the variance of $\Delta l_i - \rho \times t_i$ is smaller than that of $\Delta l_i$. Thus, we can prove that the LD 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.

### 6.2 Global Social Welfare Loss Based Discount Allocation

In fact, the LD loss based discount allocation strategy only concerns the individual loss without considering global performances (i.e., global social welfare). In other words, the passengers with large LD 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 of other passengers may be increased drastically. In other words, the difference between others' overall social welfare that includes and excludes this passenger is also an important factor in the discount allocation procedure. We call this value "global social welfare loss". Actually, this idea is motivated from the payment determination method used in Vickrey-Clarke-Groves (VCG) auction mechanism [28]. Therefore, the discount allocated to a passenger depends on not only its LD loss but also its global social welfare (SW) loss. In this subsection, we will introduce the global SW loss based discount allocation strategy, which is applied to both passengers and drivers.

The global SW loss of a passenger is evaluated by this passenger's impact on the social welfare of other passengers. It is the decrease of others' total social welfare value after excluding the passenger from the passenger set. 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 his/her global SW loss. For simplicity, let $C_i$ denote the global SW loss of $p_i \in \mathcal{P}'$. To acquire the value of $C_i$, we need to compute the other overall social welfare based on all passengers, which is denoted as $W$. Note that here the total other social welfare does not contain $p_i$. Then, we let $W_{-p_i}$ denote the social welfare based on the passengers excluding $p_i \in \mathcal{P}'$. Note that here $W_{-p_i} \geqslant W$ for $\forall p_i \in \mathcal{P}'$. This is because the system would have more assignment choices for other passengers, resulting in the increase of other overall social welfare. According to this, we can calculate the global SW loss of $p_i$ as $C_i = W_{-p_i} - W$. Since the discount of a passenger $p_i \in \mathcal{P}'$ is proportional to its SW loss, we have:

$$t_i = \frac{(\gamma \times T) \times C_i}{\sum_{p_k \in \mathcal{P}'} C_k}. \tag{7}$$

Furthermore, the global SW loss 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 global SW loss for one driver (i.e., $d_j$), denoted as $G_j$, can be calculated as the process for passengers. The original other social welfare based on all drivers is denoted as $W$, and we let $W_{-t_j}$ denote the social welfare value based on the driver set that excludes

$d_j \in \mathcal{D}$. Similarly, we use $G_j = W_{-t_j} - W$ to denote the global SW loss 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:

$$t_j = \frac{(1-\gamma) \times \kappa \times T \times G_j}{\sum_{d_x \in \mathcal{D}} G_x}. \tag{8}$$

### 6.3 Joint Discount Allocation

By combining the LD loss based strategy and the global SW loss based strategy, we propose a new discount allocation strategy for passengers, called the 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 LD and SW loss, i.e.,

$$t_i = \gamma \times T \times (\lambda \frac{\Delta l_i}{\sum_{p_k \in \mathcal{P}'} \Delta l_k} + (1-\lambda) \frac{C_i}{\sum_{p_k \in \mathcal{P}'} C_k}).$$

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

### 6.4 Detailed Discount Allocation Algorithms

Based on the above strategies, we design the discount allocation algorithms for passengers and drivers, respectively, as shown in Algorithm 3 and Algorithm 4 respectively. First, we introduce Algorithm 3. We initialize the discount values for all passengers in step 1. In step 2, the platform computes other overall social welfare values except passenger $p_i$. Then, each passenger (i.e., $p_i$) calculates its LD loss and sends it to the platform in step 3 and step 4. At the same time, the platform re-conducts the matching between drivers and passengers excluding $p_i$ and then gets a new other social welfare value in step 5. Based on this, the global SW loss of 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 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 4. Similar to Algorithm 3, in step 1, we first initialize the algorithm. In step 2, the platform computes other (except $d_j$) total 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 global SW loss of each driver in step 5, we obtain the corresponding discount according to (8), in steps 6–7. At last, we output the results in step 8.

**Algorithm 3.** Discount Allocation Algorithm for Passengers

**Input:** $\mathcal{P}'$, $\mathcal{D}$, $\gamma$, $\kappa$, $\lambda$, $T$
**Output:** $t_i$ for $\forall p_i \in \mathcal{P}'$
1: Initialization: $t_i = 0$ for $\forall p_i \in \mathcal{P}'$
2: **for** $p_i \in \mathcal{P}'$ **do**
3: Platform computes the total other social welfare except $p_i$, i.e., $W$;
4: $p_i$ calculates its LD loss, i.e., $\Delta l_i$, and then sends the value to the system platform;
5: Platform re-matches passengers $\mathcal{P}'/\{p_i\}$ and drivers $\mathcal{D}$, and computes the new other social welfare, i.e., $W_{-p_i}$;
6: Platform computes the SW loss, i.e., $C_i = W_{-p_i} - W$;
7: **for** $p_i \in \mathcal{P}'$ **do**
8: Platform calculates the discount for $p_i$, denoted as $t_i$,

$$t_i = \begin{cases} \frac{\gamma T \Delta l_i}{\sum_{p_k \in \mathcal{P}'} \Delta l_k}, & (6)\text{: LD loss,} \\ \frac{\gamma T C_i}{\sum_{p_k \in \mathcal{P}'} C_k}, & (7)\text{: global SW loss,} \\ \gamma T (\frac{\lambda \Delta l_i}{\sum_{p_k \in \mathcal{P}'} \Delta l_k} + \frac{(1-\lambda) C_i}{\sum_{p_k \in \mathcal{P}'} C_k}). \end{cases}$$

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

**Algorithm 4.** Discount Allocation Algorithm for Drivers

**Input:** $\mathcal{P}'$, $\mathcal{D}$, $\gamma$, $\kappa$, $T$
**Output:** $t_j$ for $\forall d_j \in \mathcal{D}$
1: Initialization: $t_j = 0$ for $\forall d_j \in \mathcal{D}$
2: **for** $d_j \in \mathcal{D}$ **do**
3: Platform computes the total other social welfare except the driver $d_j$, i.e., $W$;
4: Platform re-matches passengers $\mathcal{P}'$ and drivers $\mathcal{D}/\{d_j\}$, and computes the new social welfare, i.e., $W_{-d_j}$;
5: Platform computes the global SW loss $G_j = W_{-d_j} - W$;
6: **for** $d_j \in \mathcal{D}$ **do**
7: Platform calculates the discount for $d_j$ based on (8);
8: **return** $t_j$ for $\forall d_j \in \mathcal{D}$;

Our discount allocation mechanisms would not leak the location privacy of passengers. For the LD loss based discount allocation, each passenger reports the loss to the SP. Even if we assume the loss of each passenger is proportional to the extra pick-up distance, the SP cannot reveal the passenger's location. It is because the passenger would evaluate the loss by comparing the distance from the nearest driver, which is not known by the SP. Without knowing which driver is the nearest to

the passenger, it is hard for the SP to infer the passenger's actual location based on the loss. For the global SW loss based strategy, the discount is calculated by using the VCG auction mechanism. The discount price is not proportional to the extra pick-up distance, but is mainly determined by the decrease of others' overall social welfare value after excluding the passenger from the passenger set. It is difficult for the SP to infer the passenger's location based on the difference in social welfare. The joint discount allocation strategy does not reveal more information to SP other than the LD loss and the global SW loss. Besides, the SP could not infer the nearest driver to each passenger based on the global SW loss. Thus, the joint discount allocation strategy would not leak passengers' location privacy.

### 6.5 Example

The LD loss based discount allocation strategy is easy to understand. To better understand the global SW loss based discount allocation strategy, we use an example shown in Fig.7 to illustrate the allocation procedure. We suppose three passengers and four drivers in the system. The distance values between drivers and passengers are shown as follows.

$dis(p_1, d_1) = 1$, $dis(p_1, d_2) = 3$, $dis(p_1, d_3) = 4$,
$dis(p_1, d_4) = 3$, $dis(p_2, d_1) = 3$, $dis(p_2, d_2) = 4$,
$dis(p_2, d_3) = 5$, $dis(p_2, d_4) = 7$, $dis(p_3, d_1) = 6$,
$dis(p_3, d_2) = 9$, $dis(p_3, d_3) = 7$, $dis(p_3, d_4) = 8$.

In the example, we directly use the distance instead of the social welfare value. Based on this given information, the platform can output the optimal matching between passengers and drivers so that the overall pick-up distance can be minimized (i.e., the social welfare is maximized), as shown in Fig.7(a). Then, the discount allocation procedure is conducted as follows.

Note that since no passenger is assigned to the driver $d_4$, he/she has no SW loss for other drivers. Next, we calculate the global SW loss for drivers $d_1$, $d_2$, and $d_3$. First, for the driver $d_1$, we first exclude $d_1$ and conduct the matching between the passengers and the remaining drivers. In order to minimize the total distance, we get the matching results in Fig.7(b). Now, we compute the total pick-up distance of other drivers except $d_1$, i.e., $4+7+3=14$. In the initial matching, the total distance of $d_2$, $d_3$ and $d_4$ is $4+7+0=11$. Therefore, the difference is calculated as $14-11=3$. That is, the global SW loss of $d_1$ is 3. Similarly, after excluding $d_2$ and $d_3$, we conduct the same operations. The corresponding results are shown in Fig.7(c) and Fig.7(d), respectively. At the same time, the global SW loss for $d_2$ and $d_3$ is calculated as in Fig.7.

After determining the SW loss for each driver, the allocated discount is computed as follows. Note that the total discounts for all drivers are $(1-\gamma)\times\kappa\times T$. Therefore we just need to focus on how to calculate the percentages for each driver. According to the proposed discount allocation algorithm, we get that the percentages for $d_1$, $d_2$, $d_3$ and $d_4$ are $\frac{3}{16}$, $\frac{5}{16}$, $\frac{1}{2}$ and 0, respectively. Further, the discounts allocated to $d_1$, $d_2$, $d_3$ and $d_4$ are $t_1=\frac{3}{16}(1-\gamma)\times\kappa\times T$, $t_2=\frac{5}{16}(1-\gamma)\times\kappa\times T$, $t_3=\frac{1}{2}(1-\gamma)\times\kappa\times T$ and $t_4=0$. Moreover, we can also calculate the discount values for all passengers based on (7), by using the global SW loss based discount allocation strategy. The procedure is the same as the discount allocation for drivers in the example, thereby we will omit the detailed calculation here.

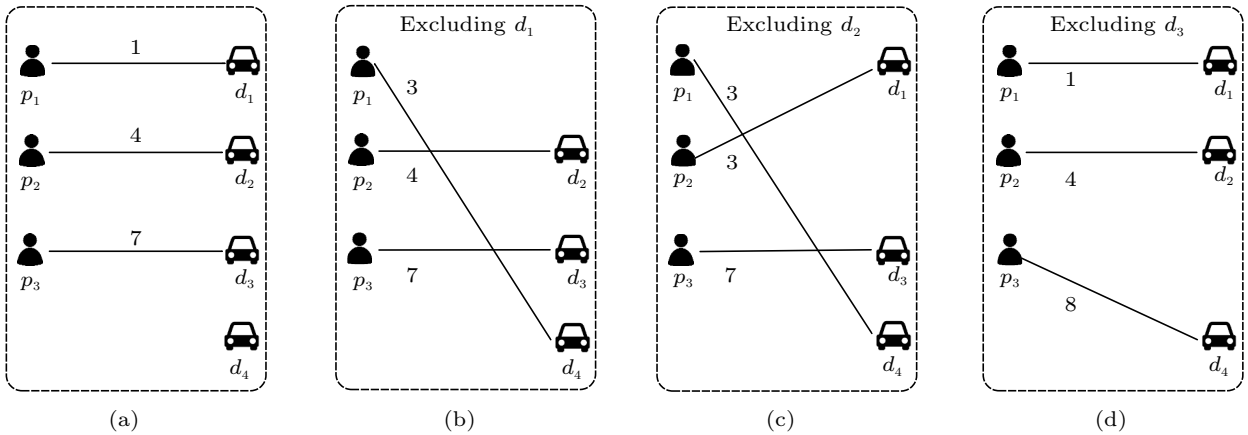


Fig.7. Walk-through example of the global social welfare loss based discount allocation. (a) Initial matching between passengers and drivers. (b) The SW loss of $d_1$ is $(4+7+3)-(4+7+0)=3$. (c) The SW loss of $d_2$ is $(3+7+3)-(1+7+0)=5$. (d) The SW loss of $d_3$ is $(1+4+8)-(1+4+0)=8$.

## 7 Experiment

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

### 7.1 Experimental 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 using uniform distribution. Specifically, we first generate a planar size in $30 \times 30$ km$^2$. 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 randomly sampled from the order and trace data① in Chengdu, China, from Didi Inc. We mark part of the locations in the map of Chengdu, as shown in Fig.8. To the best of our knowledge, there is no available dataset that contains privacy requirements of passengers. Therefore, we assume that the privacy requirements obey normal distribution. The mean of the distribution is denoted as $\mu$, and the standard deviation is set as $\mu/3$, which could guarantee that 99.7% of the generated privacy requirements are 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 as the same as that of passengers.

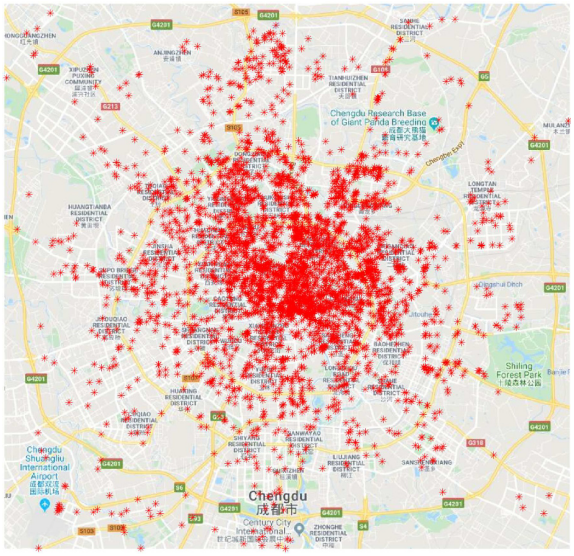


Fig.8. Illustration of the real-world dataset.

We first investigate the matching performance of our scheme on both datasets. We compare our scheme with the simple spatial cloaking approach in which each passenger greedily chooses the nearest available driver. The comparison algorithm is denoted as greedy. We also compare our scheme with the optimal matching in which the actual locations of passengers are used. Then, we simulate the discount allocation strategies. Note that the location information of drivers is public to the passengers, thereby the loss of each passenger is calculated easily. Then, the global SW loss based allocations for drivers and passengers are computed according to two matching results. In the joint discount allocation algorithm, we set the parameter $\lambda$ from $\{0.3, 0.5, 0.7\}$, and the default $\lambda$ is 0.5. In addition, the total discounts allocated to drivers or passengers are proportional to the number of drivers or passengers.

### 7.2 Simulation Results

Fig.9 shows the comparison of different algorithms on the overall pick-up distances. Fig.9(a) illustrates the 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 the blue lines shown in the figure. By comparing the red line and the blue line, we can verify that larger privacy requirements would result in larger overall pick-up distance. Fig.9(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 $\mu$ from 1 to 2, the relative difference between the red line and the blue line is larger in Fig.9(b) than that in Fig.9(a). Although the effect of the value of $\mu$ is more obvious, our scheme still outperforms the simple spatial cloaking approach.

Fig.10 shows the comparison of different algorithms on the pick-up distance distribution. Figs.10(a) and 10(b) plot the cumulative distribution function ($F(x)$)

①https://gaia.didichuxing.com, Apr. 2020.

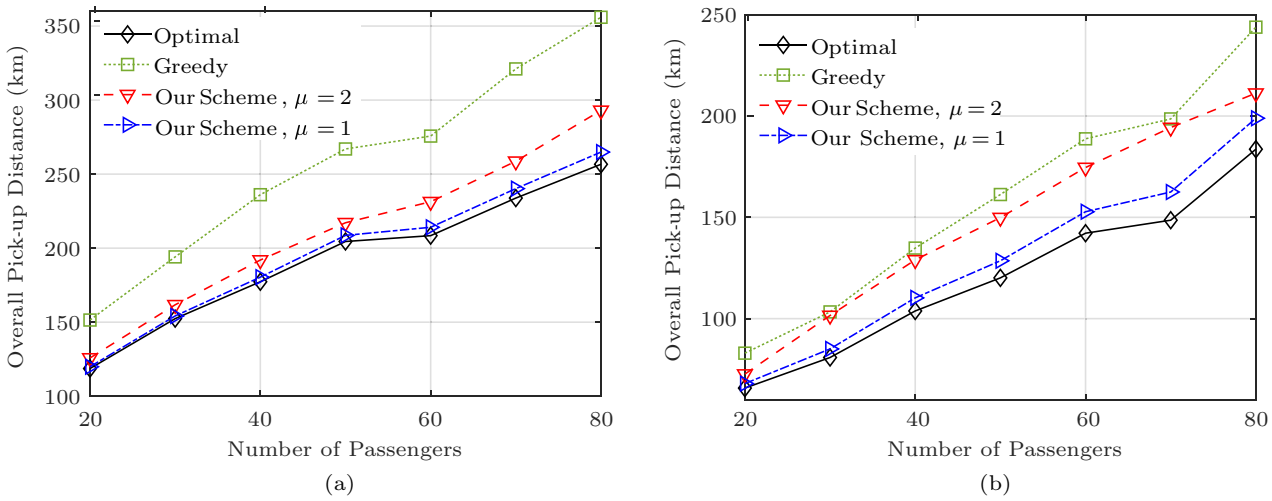


Fig.9. Comparison of overall pick-up distances [12]. (a) On the synthetic dataset. (b) On the real-world dataset.

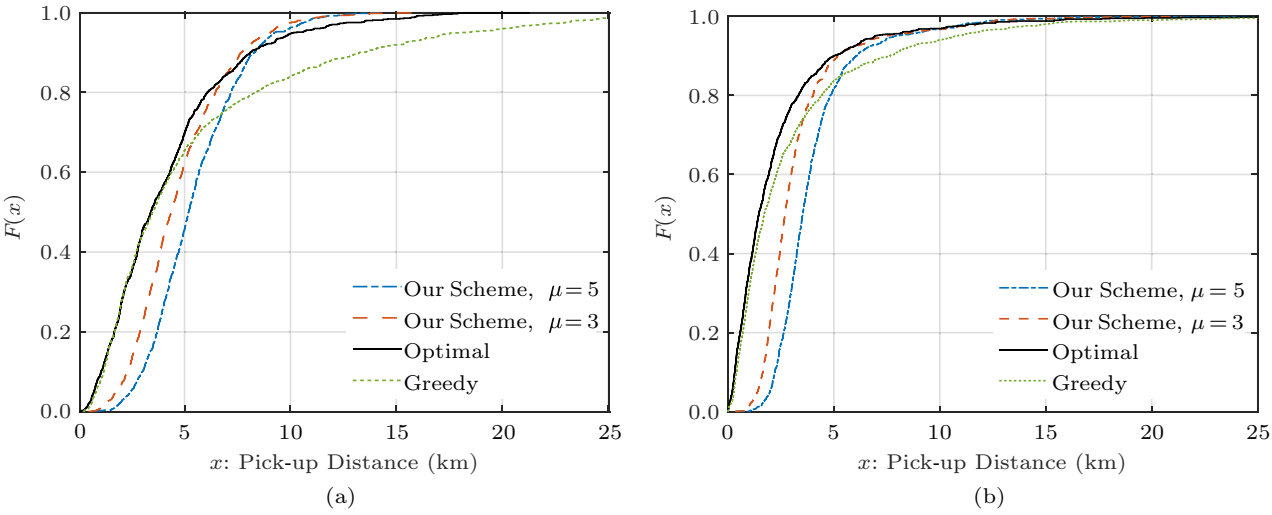


Fig.10. Comparison of pick-up distance distributions [12]. (a) On the synthetic dataset. (b) On the real-world dataset.

of the pick-up distances ($x$). From Fig.10(a), we can find out that 60% of the pick-up distances are less than 4.4 km when the simple spatial cloaking approach is used. The corresponding value of our scheme is 4.9 km when $\mu = 3$, which is larger. In contrast, when investigating 80% of the pick-up distances, they are less than 8.4 km when the simple spatial cloaking approach is used, and the corresponding value of our scheme is 6.4 km when $\mu = 3$. This shows that allowing passengers to choose the nearest driver could benefit some passengers whose pick-up distance is relatively small, while it also has negative effects on the pick-up distances of some other passengers. With our scheme, the pick-up distances are more concentrated compared with the simple spatial cloaking approach. We can find out the similar conclusion in Fig.10(b).

We test our extended scheme, which considers the distance constraint set by drivers, on the real-world dataset. We modify the optimal matching and greedy algorithm to fit in the distance constraint. For the optimal matching which is labeled as "optimal" in Fig.11, the actual locations of passengers are known. Therefore, the actual pick-up distances between drivers and passengers could be calculated. The edges which break the distance constraint are deleted in the bipartite graph before the weighted bipartite matching is applied. For the greedy approach, passengers would choose the nearest driver whose pick-up distance limitation is not violated. The simulation result is shown in Fig.11.

In Fig.11, we compare the average pick-up distances and the number of unserved passengers of different schemes. As we explained in Section 5, allowing drivers to set pick-up distance limitations might result in unserved passengers. Especially, in our experiment, the

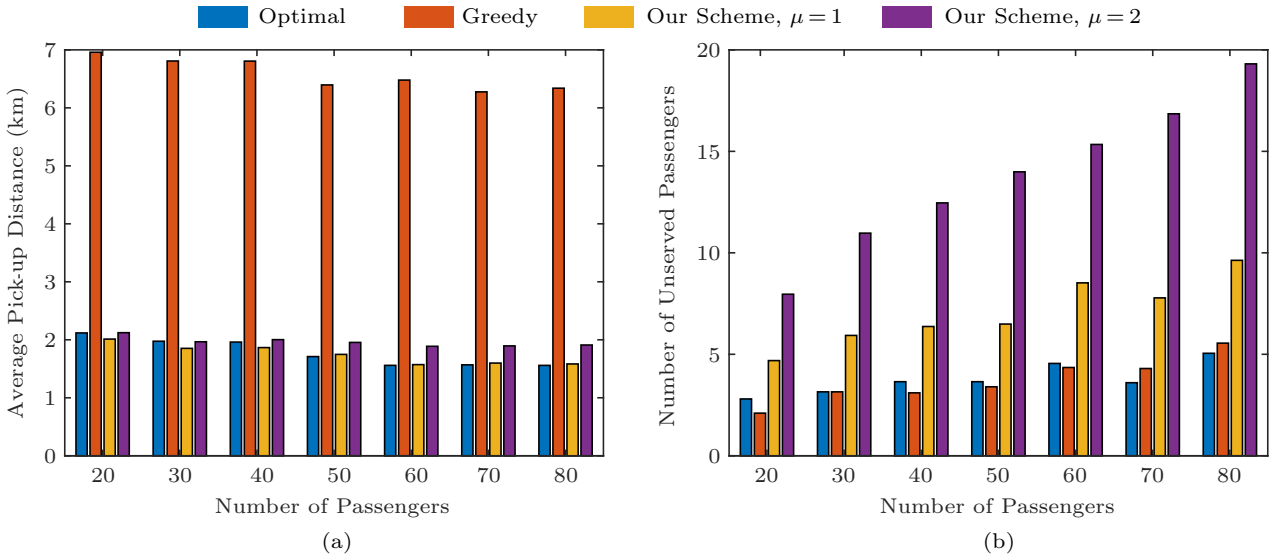


Fig.11. Performance of extended scheme on the real-world dataset. (a) Average pick-up distances. (b) Number of unserved passengers.

number of drivers is the same with the number of passengers. In the experiment, the maximum pick-up distances of drivers are sampled from the normal distribution $\mathcal{N}(5, 5/3)$, i.e., the mean is 5 and the standard deviation is $5/3$, and the negative limitations are ignored. Fig.11(a) shows the average pick-up distances of passengers who are assigned drivers to serve. We find out that the matching-based schemes could achieve shorter pick-up distances than the greedy approach. One counterintuitive result is that the average pick-up distance of the optimal could be larger than that of our scheme. The reason is that the number of unserved passengers is different. Although the optimal approach has larger average pick-up distances, it has a smaller number of unserved passengers. Fig.11(b) shows the number of unserved passengers. We find that the optimal approach always achieves a smaller number than our scheme. It also shows the trade-off of our extended scheme. Considering that the unserved passengers need to wait for the next round of order dispatch, the extended scheme is more suitable for the case where there are more drivers than passengers.

We evaluate the allocated discount for each passenger based on three allocation algorithms, as shown in Fig.12(a). We see that the allocated discount differences among all passengers in the LD loss based algorithm are larger than those of the other algorithms, and the discount differences in the global SW loss based

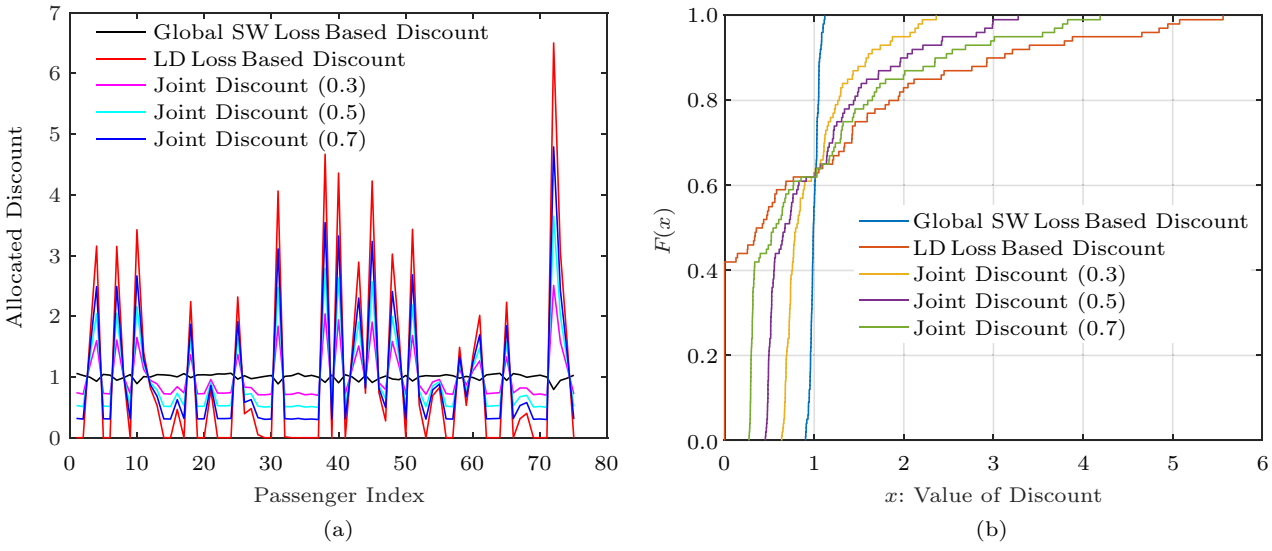


Fig.12. Discount distribution for passengers [12]. (a) Passenger index. (b) CDF (cumulative distribution function).

algorithm are the smallest. This means that the loss of each passenger dominates in the joint discount allocation algorithm. Moreover, we present the cumulative distribution function (CDF) of the allocated discount for three algorithms in Fig.12(b). We also show the discount results when changing parameter $\lambda$. We find that in the LD 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 global SW loss based algorithm, the differences in the allocated discounts for passengers are small. When we change parameter $\lambda$, all passengers will get certain discounts.

We also evaluate the influences of the individual privacy requirement, as shown in Fig.13. We let one individual privacy radius change from 1 km to 15 km, while keeping the others unchanged. Moreover, to evaluate the effect of the number of passengers, we set the number of passengers and drivers as $\{25, 50, 75\}$. In this setting, we observe that the individual local distance loss for the certain passenger would increase along with the increase of his/her individual privacy requirement, as shown in Fig.13(a). And we find that the more the passengers are involved, the smaller the local distance loss one individual would suffer. This is because the density of passengers and drivers will increase. Accordingly, the matching distance between passengers and drivers would decrease. Thus, the local distance loss for each passenger will decrease. On the other hand, we evaluate different fares on the privacy requirement, as shown in Fig.13(b). When one passenger increases his/her privacy requirement, the privacy fare, which is proportional to his/her clocking region, would increase accordingly. Meanwhile, the LD loss based discount will also increase according to the former conclusion. Then, the additional fare (which denotes the difference between the privacy fare and the allocated discount) has an upward trend. These observations are consistent with our theoretical analysis.

Furthermore, we show the variance of the additional expense in Definition 1 in Fig.14(a). We find that the LD loss based algorithm always achieves the minimum variance value while the global SW loss 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.14(b). Since the true locations of passengers are not invisible to drivers, only the global SW loss based algorithm can be applied for drivers. We thus find that the allocated discount values for each passenger change a little.

## 8 Conclusions

A cloaking region based privacy-preserving order dispatch system for ride-hailing services was investigated in this paper. Unlike the previous schemes in which passengers choose the desired drivers, we investigated the approach that lets the service provider match passengers and drivers in a centralized way. Our ap-

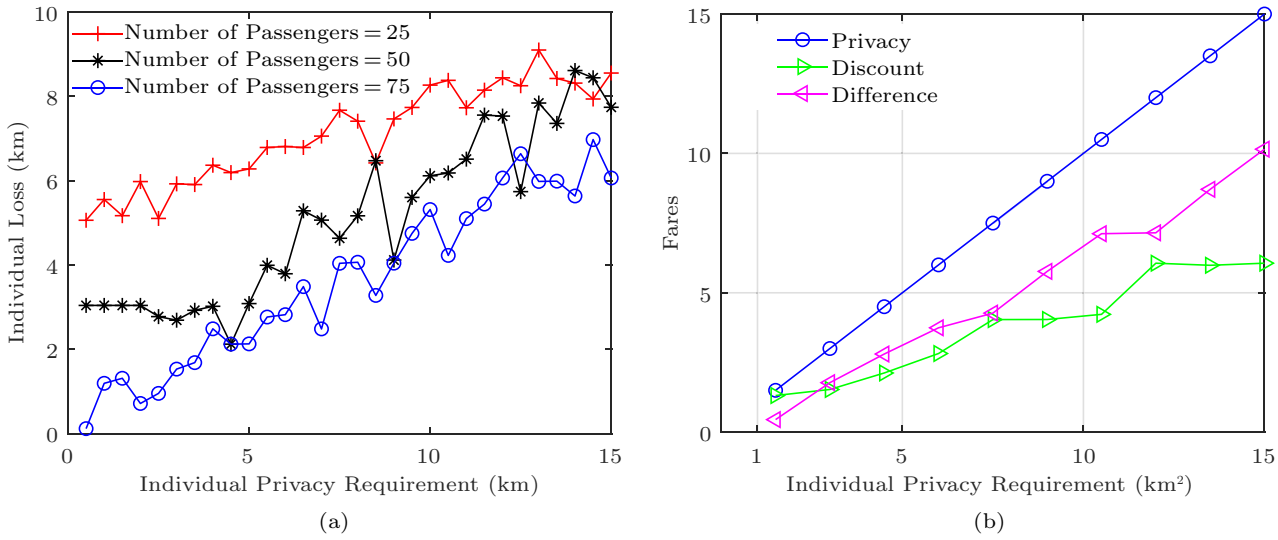


Fig.13. Influences of individual privacy requirement. (a) Privacy vs LD loss. (b) On different fares.

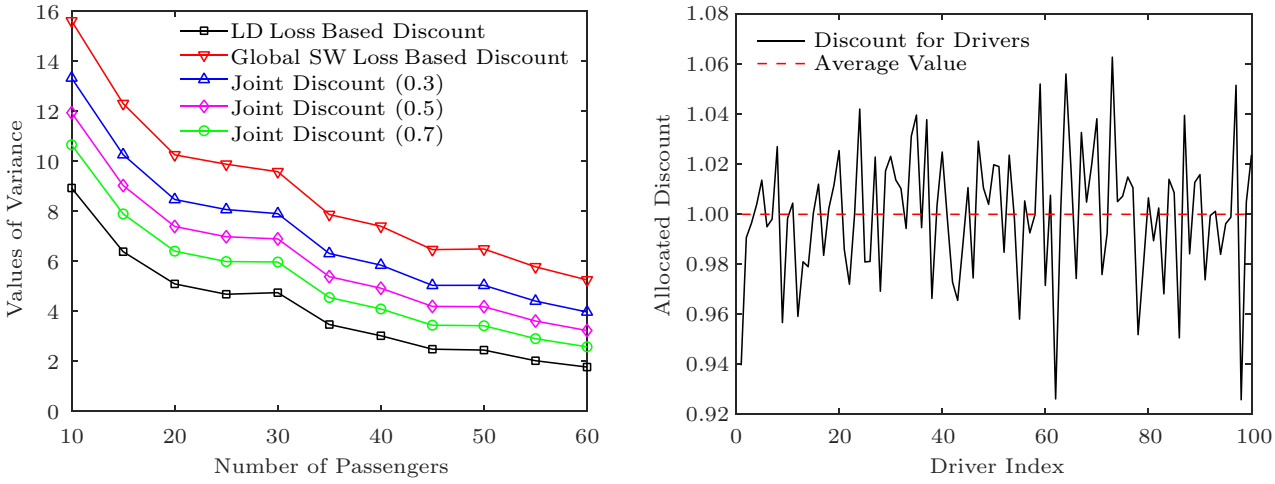


Fig.14. Discount distribution for passengers and drivers [12]. (a) Simulation of fairness. (b) Discount for drivers.

proach could efficiently prevent the privacy inference attacks proposed in [10]. In addition, our approach would not introduce large communication overhead. Based on the matching approach, we proposed to maximize the social welfare (i.e., minimizing the overall pick-up distance) of the order dispatch. Although our scheme could not achieve the optimal social welfare since the actual passenger locations are unknown to the service provider, we could guarantee that the overall pick-up distance of our matching result is upper bounded by $OPT + \sqrt{2}\sum_{p_i \in \mathcal{P}} \sqrt{S_i}$. Also, the strong privacy defined in [10] could be guaranteed. Besides considering passengers' interest, we allow drivers to set up their maximum pick-up distances in our extended scheme. The extended scheme is suitable when the number of drivers is larger than the number of passengers. As for discount allocation, we introduced three strategies to make up for the loss of all individuals caused by global matching. Finally, we conducted lots of experiments based on both synthetic and real-world datasets to verify the performance of our scheme.

## References


[1] Duan Y, Mosharraf T, Wu J, Zheng H. Optimizing carpool scheduling algorithm through partition merging. In *Proc. the IEEE Int. Conf. Communication*, May 2018.

[2] Gao G, Xiao M, Zhao Z. Optimal multi-taxi dispatch for mobile taxi-hailing systems. In *Proc. the 45th IEEE Int. Conf. Parallel Processing*, August 2016, pp.294-303.

[3] Shokri R, Theodorakopoulos G, le Boudec J Y, Hubaux J P. Quantifying location privacy. In *Proc. the 32nd IEEE Symp. Security and Privacy*, May 2011, pp.247-262.

[4] Meyer-Lee G, Shang J, Wu J. Location-leaking through network traffic in mobile augmented reality applications. In *Proc. the 37th IEEE Int. Performance Computing and Communications Conf.*, November 2018.

[5] Damiani M L, Bertino E, Silvestri C. The PROBE framework for the personalized cloaking of private locations. *Trans. Data Privacy*, 2010, 3(2): 123-148.

[6] Xue M, Kalnis P, Pung H K. Location diversity: Enhanced privacy protection in location based services. In *Proc. the 4th Int. Symp. Location and Context Awareness*, May 2009, pp.70-87.

[7] Pham A, Dacosta I, Endignoux G, Troncoso-Pastoriza J R, Huguenin K, Hubaux J P. ORide: A privacy-preserving yet accountable ride-hailing service. In *Proc. the 26th USENIX Security Symp.*, August 2017, pp.1235-1252.

[8] Aïvodji U M, Huguenin K, Huguet M J, Killijian M O. SRide: A privacy-preserving ridesharing system. In *Proc. the 11th ACM Conf. Security & Privacy in Wireless and Mobile Networks*, June 2018, pp.40-50.

[9] He Y, Ni J, Wang X, Niu B, Li F, Shen X. Privacy-preserving partner selection for ride-sharing services. *IEEE Trans. Vehicular Technology*, 2018, 67(7): 5994-6005.

[10] Khazbak Y, Fan J, Zhu S, Cao G. Preserving location privacy in ride-hailing service. In *Proc. the 2008 IEEE Conf. Communications and Network Security*, May 2018.

[11] Aurenhammer F. Voronoi diagrams — A survey of a fundamental geometric data structure. *ACM Computing Surveys*, 1991, 23(3): 345-405.

[12] Duan Y, Gao G, Xiao M, Wu J. A privacy-preserving order dispatch scheme for ride-hailing services. In *Proc. the 16th Int. Conf. Mobile Ad-hoc and Smart Systems*, November 2019.

[13] Hadiwardoyo S A, Patra S, Calafate C T, Cano J C, Manzoni P. An intelligent transportation system application for smartphones based on vehicle position advertising and route sharing in vehicular ad-hoc networks. *Journal of Computer Sci. and Tech.*, 2018, 33(2): 249-262.

[14] Beresford A R, Stajano F. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2003, 2(1): 46-55.

[15] Liu A, Li Z X, Liu G F, Zheng K, Zhang M, Li Q, Zhang X. Privacy-preserving task assignment in spatial crowdsourcing. *Journal of Computer Sci. and Tech.*, 2017, 32(5): 905-918.

[16] Hoh B, Gruteser M. Protecting location privacy through path confusion. In *Proc. the 1st IEEE Int. Conf. Security and Privacy for Emerging Areas in Communications Networks*, September 2005, pp.194-205.
[17] Sánchez D, Martínez S, Domingo-Ferrer J. Co-utile P2P ridesharing via decentralization and reputation management. *Transportation Research Part C: Emerging Technologies*, 2016, 73: 147-166.
[18] Goel P, Kulik L, Ramamohanarao K. Optimal pick up point selection for effective ride sharing. *IEEE Trans. Big Data*, 2017, 3(2): 154-168.
[19] Dai C, Yuan X, Wang C. Privacy-preserving ridesharing recommendation in geosocial networks. In *Proc. the 5th Int. Conf. Computational Social Networks*, August 2016, pp.193-205.
[20] Aïvodji U M, Gambs S, Huguet M J, Killijian M O. Meeting points in ridesharing: A privacy-preserving approach. *Transportation Research Part C*: *Emerging Technologies*, 2016, 72: 239-253.
[21] Li H, Zhu H, Du S, Liang X, Shen X S. Privacy leakage of location sharing in mobile social networks: Attacks and defense. *IEEE Trans. Dependable and Secure Computing*, 2018, 15(4): 646-660.
[22] Zhang N, Zhong S, Tian L. Using blockchain to protect personal privacy in the scenario of online taxi-hailing. *Int. Journal of Computers Communications Control*, 2017, 12(6): 886-902.
[23] Pham A, Dacosta I, Jacot-Guillarmod B, Huguenin K, Hajar T, Tramèr F, Gligor V, Hubaux J P. PrivateRide: A privacy-enhanced ride-hailing service. *Proceedings on Privacy Enhancing Technologies*, 2017, 2017(2): 38-56.
[24] Liang X, Li X, Lu R, Lin X, Shen X. UDP: Usage-based dynamic pricing with privacy preservation for smart grid. *IEEE Trans. Smart Grid*, 2013, 4(1): 141-150.
[25] Zhuo X, Gao W, Cao G, Dai Y. Win-coupon: An incentive framework for 3G traffic offloading. In *Proc. the 19th Annual IEEE Int. Conf. Network Protocols*, October 2011, pp.206-215.
[26] Gao G, Xiao M, Wu J, Huang L, Hu C. Truthful incentive mechanism for nondeterministic crowdsensing with vehicles. *IEEE Trans. Mobile Computing*, 2018, 17(12): 2982-2997.
[27] Edelman B, Ostrovsky M, Schwarz M. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American Economic Review*, 2007, 97(1): 242-259.
[28] Vickrey W. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 1961, 16(1): 8-37.
[29] Cheng R, Zhang Y, Bertino E, Prabhakar S. Preserving user location privacy in mobile data management infrastructures. In *Proc. the 6th Int. Workshop on Privacy Enhancing Technologies*, June 2006, pp.393-412.

**Yubin Duan** received his B.S. degree in mathematics and physics from University of Electronic Science and Technology of China, Chengdu, in 2017. He is currently a Ph.D. candidate in the Department of Computer and Information Sciences, Temple University, Philadelphia. His current research focuses on urban computing.

**Guo-Ju Gao** received his B.S. degree in information security from the University of Science and Technology of Beijing, Beijing, in 2014. He is currently working toward his Ph.D. degree in computer science and technology in the University of Science and Technology of China, Hefei. His research interests include privacy preservation, mobile crowdsensing and incentive mechanisms.

**Ming-Jun Xiao** is a professor in the School of Computer Science and Technology at the University of Science and Technology of China (USTC), Hefei. He received his Ph.D. degree in computer science and technology from USTC in 2004. His research interests include crowdsourcing, mobile social networks, mobile cloud computing, blockchain, data security and privacy. He has published more over 80 papers in referred journals and conferences, including TMC, TC, TPDS, TKDE, TSC, etc. He served as the TPC member of INFOCOM 2020, DASFAA 2020, INFOCOM 2019, ICDCS 2019, DASFAA 2019, etc. He is on the reviewer board of several top journals such as TMC, TON, and so on.

**Jie Wu** is the director of the Center for Networked Computing and a Laura H. Carnell professor at Temple University, Pennsylvania. He also serves as the director of International Affairs at College of Science and Technology. Prior to joining Temple University, he was a program director at the National Science Foundation and was a distinguished professor at Florida Atlantic University. His current research interests include mobile computing and wireless networks, routing protocols, cloud and green computing, network trust and security, and social network applications. Dr. Wu regularly publishes in scholarly journals, conference proceedings, and books. He serves on several editorial boards, including IEEE Transactions on Service Computing and the Journal of Parallel and Distributed Computing. Dr. Wu was general co-chair for IEEE MASS 2006, IEEE IPDPS 2008, IEEE ICDCS 2013, ACM MobiHoc 2014, ICPP 2016, and IEEE CNS 2016, and program co-chair for IEEE INFOCOM 2011 and CCF CNCC 2013. He was an IEEE Computer Society Distinguished Visitor, ACM Distinguished Speaker, and chair for the IEEE Technical Committee on Distributed Processing (TCDP). Dr. Wu is a CCF Distinguished Speaker and a fellow of IEEE. He is the recipient of the 2011 CCF Overseas Outstanding Achievement Award.