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

Yubin Duan, Guojun Gao, Mingjun Xiao, and Jie Wu
Dept. of Computer and Info. Sciences
Temple University
1. Introduction

- **Ride-hailing system**
  - *Service Provider* (SP): Uber and Didi
  - *Order dispatch*: matching *passengers* (P) and *drivers* (D)

- **Privacy concerns**
  - Passenger locations are exposed to the SP
  - SP could infer passengers’ habits \(^{[1]}\)

---

\(^{[1]}\) Shokri, R., Theodorakopoulos, G., Le Boudec, J. Y., and Hubaux, J. P., Quantifying Location Privacy (IEEE SSP '11)
Motivation

- **Cloaking region** $S_i$ (for privacy protection)
  - Passenger $p_i$ sends a fake location $p_i'$ to SP
  - SP cannot infer passenger’s exact location in $S_i$

- **How to perform order dispatch** (for different $S_i$)?
  - Let passengers choose the nearest driver [2], or
  - Let SP match in a centralized manner (this paper)

Privacy Attack

- **Passenger choosing** [2]

- **Attack model** [2]
  - Voronoi graph:
    - nearest driver
    - passenger location
    - driver locations

Preventing Privacy Attack

- A probabilistic mechanism \[^{[2]}\]
  - Form and sort driver set \(D\) with \(k\) nearest drivers
  - Partition \(D\) into \(D_1\) and \(D_2\) based on distance
  - Pick a driver from \(D_1\) (\(D_2\)) with a higher (lower) probability

- Guarantee privacy (based on prior probabilities) \[^{[2]}\]
  - Problem: not optimize pick-up distances, locally nor globally

Our Approach

- Optimize social welfare
  - Minimize the total pick-up distance (bipartite matching)

- Performance loss
  - Travel fares + privacy fares - discount

Pick-up distance by matching: blue (based on p’) > red (based on p)
2. Social Welfare Optimization

- Social welfare: \(- dis(p_i, d_j)\) (negation of pick-up distance)
- Privacy requirement: \(|S_i|\)

\[
\begin{align*}
\text{max} \quad W &= -x_{ij} \text{dis}(p_i, d_j) \\
\text{s.t.} \quad \sum_{d_j} x_{ij} &= 1, \quad x_{ij} \in \{0, 1\}, \forall p_i \\
\sum_{p_i} x_{ij} &\leq 1, \quad x_{ij} \in \{0, 1\}, \forall d_j \\
\|p_i - p_i'\|_\infty &\leq \sqrt{S_i}/2, \forall p_i
\end{align*}
\]

Maximize social welfare

All passengers matched

Not all drivers matched

Privacy constraint
**Theorem:** actual pick-up distance

\[ \sum_{p_i} \text{blue} \leq \sum_{p_i} (\text{red} + \sqrt{2S_i}) \]

- **Proof sketch**

  **Optimality of bipartite matching:**

  \[ \sum_{p_i} \text{black} \leq \sum_{p_i} \text{green} \]

  **Triangle inequality:**

  \[ \sum_{p_i} \text{blue} \leq \sum_{p_i} (\text{black} + \text{grey}) \]

  \[ \sum_{p_i} \text{green} \leq \sum_{p_i} (\text{red} + \text{grey}) \]

  **Combining:**

  \[ \sum_{p_i} \text{blue} \leq \sum_{p_i} (\text{red} + 2\text{grey}) \]
3. Discount Allocation

- **Profit distribution**
  - SP
  - Drivers in D (global)
  - Passengers in P (local + global)

- **Local distance loss** (for P)
  - The difference between actual pick-up and nearest distance
  - $p_3$ local loss: blue line - yellow line (nearest)

- **Global social welfare loss** (for a party in P or D)
  - The difference between others’ social welfare that includes and excludes this party \(^3\)

---

\(^3\) Krishna, V. and Motty, P., Efficient mechanism design (Available at SSRN 64934, 1998).
Global Social Welfare Loss

Global social welfare (SW) loss for passengers/drivers based on VCG [3]

<table>
<thead>
<tr>
<th>Passengers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>p_1</td>
<td>2/5</td>
</tr>
<tr>
<td>p_2</td>
<td>1/5</td>
</tr>
<tr>
<td>p_3</td>
<td>2/5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drivers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>d_1</td>
<td>3/5</td>
</tr>
<tr>
<td>d_2</td>
<td>1/5</td>
</tr>
<tr>
<td>d_3</td>
<td>1/5</td>
</tr>
</tbody>
</table>

social welfare for \(p_2\) and \(p_3\) is 
\[-(5+8) = -13\]

social welfare for \(p_2\) and \(p_3\) is 
\[-(4+7) = -11\]

the SW loss of \(p_2\) is 
\[-(5+6)-(-(4+8)) = 1\]

the SW loss of \(p_3\) is 
\[-(3+4)-(-(4+5)) = 2\]

the SW loss of \(p_1\) is the difference, i.e., 
\[-11-(-13) = 2\]

Discount Allocation Strategy

- For drivers in D
  - Discount is based on global social welfare (SW) loss

- For passengers in P
  1. Discount is based on global SW loss;
  2. Discount is based on local distance (LD) loss;
  3. Combine 1) and 2), i.e.,

\[ \lambda \times LD + (1 - \lambda) \times SW \]
4. Experiment

- Synthetic and real-world dataset
  - Synthetic: $p_i, d_j$ (uniform distribution)
  - Real-world (Didi passenger dataset):
    - $p_i$: Didi trace data in Chengdu; $d_j$: uniform distribution
  - Privacy settings: $S_i \sim \mathcal{N}(\mu, \mu/3)$ (normal distribution)

- Dataset statistics

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Didi’s trajectory data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Span</td>
<td>11/1/2016 - 11/30/2016</td>
</tr>
<tr>
<td>Number of orders</td>
<td>691,269</td>
</tr>
</tbody>
</table>

[4] Identification of urban regions’ functions in Chengdu, China, based on vehicle trajectory data (NCBI)
Experiment Results

- Overall pick-up distance

- **Greedy**: each passenger greedily chooses the nearest driver
- **Optimal**: SP matches based on real passenger & driver locations

![Graphs showing comparison between Greedy, Optimal, and our scheme with different parameters for synthetic and Didi passenger datasets.](image-url)
Experiment Results (1)

- Impact of privacy requirement

**settings**: \( \mu = 5 \) km for other passengers with uniform distributions

**conclusion**: the higher the privacy, the more the local distance loss.

**settings**: Privacy: \(|S_i|\)  
Difference = Privacy - Discount 2

**conclusion**: the higher the privacy, the more the difference value.
Experiment Results (2)

- Evaluation on three discount allocation strategies

**settings:** number of passengers = 75, total distributed profits = 75, uniform distribution

**conclusion:** the values of global social welfare loss for all passengers are smoother than that of their local distance loss.
5. Summary

- Privacy-preserving order dispatch scheme
  - SP matches passengers and drivers with privacy requirement

- The trade-off between performance and privacy
  - Derive the bound of performance loss
  - Propose to allocate discounts to make up the loss

- Experiments on real-world/synthetic datasets
  - Show the matching performance with different settings
  - Evaluate the fares and discount with different settings
Thank you

Q & A