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

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

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

- **Privacy issues**
  - Passenger locations are exposed to the SP
  - Public concerns: 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)

---

\[2\] Khazbak, Y., Fan, J., Zhu, S. and Cao, G., Preserving location privacy in ride-hailing service (IEEE CNS’18)
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

Actual pick-up distance: blue > red
2. Social Welfare Optimization

- Social welfare: \(- \text{dis}(p_i, d_j)\) (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
Bounded Performance Loss

**Theorem:** actual pick-up distance

\[ \sum_{p_i} M_i \leq \sum_{p_i} (OPT_i + \sqrt{2S_i}) \]

- **Proof sketch**
  - Use triangle inequality and optimality of bipartite matching
3. Discount Allocation

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

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

- **Global social welfare loss** (for passengers and drivers)
  - The difference between others’ social welfare that includes and excludes this user [3]

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>p₁</td>
<td>2/5</td>
</tr>
<tr>
<td>p₂</td>
<td>1/5</td>
</tr>
<tr>
<td>p₃</td>
<td>2/5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drivers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>d₁</td>
<td>3/5</td>
</tr>
<tr>
<td>d₂</td>
<td>1/5</td>
</tr>
<tr>
<td>d₃</td>
<td>1/5</td>
</tr>
</tbody>
</table>

social welfare for p₂ and p₃ is $(-(5+8)) = -13$

social welfare for p₂ and p₃ is $(-(4+7)) = -11$

the SW loss of p₂ is $-(5+6)-(-(4+8)) = 1$

the SW loss of p₃ is $-(3+4)-(-(4+5)) = 2$

the SW loss of p₁ is the difference, i.e., $-11-(-13) = 2$

Discount Allocation Strategy

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

- For passengers
  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 overall pick-up distance vs number of passengers for Synthetic and Didi passenger datasets.](image)
Experiment Results

- 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 $= \text{Privacy} - \text{Discount 2}$

**Conclusion:** the higher the privacy, the more the difference value.
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