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

Yubin Duan, Guoju 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 \[1\].

---

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\textsuperscript{[2]}
  - Form and sort driver set D with k nearest drivers
  - Partition D into D\textsubscript{1} and D\textsubscript{2} based on distance
  - Pick a driver from D\textsubscript{1} (D\textsubscript{2}) with a higher (lower) probability

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

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

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

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 \(^3\)

---

\(^3\) 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 \[^3\]

- 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., \(-11 - (-13) = 2\)

<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>

\[^3\] Krishna, V. and Motty, P., Efficient mechanism design (Available at SSRN 64934, 1998).
Discount Allocation Strategy

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

- For passengers
  1. Discount is based on its global SW loss;
  2. Discount is based on its local distance (LD) loss;
  3. Combining 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 on real passenger and driver locations
Impact of privacy requirement

**Experiment Results**

- **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