ORide: A Privacy-Preserving yet Accountable Ride-Hailing Service

Anh Pham¹, Italo Dacosta¹, Guillaume Endignoux¹, Juan Ramón Troncoso-Pastoriza¹, Kévin Huguenin², and Jean-Pierre Hubaux¹

¹EPFL, ²UNIL
Ride-Hailing Services (RHSs)

- Millions of users, billions of rides, hundreds of cities

- In just six months, China’s Didi doubles daily rides to 20m
  By Steven Millward
  2:50 PM at Oct 26, 2016 | 1 min read

- Uber Now Has 40 Million Monthly Riders Worldwide
  By Kira Kulishova
  Oct 20, 2016
  Uber is still growing.

- Technology News | Mon Jul 18, 2016 | 11:06 am EDT
  Uber reaches 2 billion rides six months after hitting its first billion

- Uber and Lyft Are Now Bigger than Taxis and Rental Cars Combined
  By Ian Mount
  Oct 20, 2016
Privacy Problems

For riders

Lyft Sued For Violating California Privacy Law
by Wendy Davis @wendyndavis, February 12, 2014

Uber employees used app to ‘spy on ex-lovers, politicians and celebrities’
Thousands of staff have access to ‘God View’ mode which can be used to track customers’ movements, former security executive alleges
by Ben Chapman - Tuesday 12 December 2016 12:35 GMT | 230 comments

For drivers

TECH & SCIENCE
UBER DRIVERS HAVE PRIVACY PROBLEMS TOO
BY POLLY HOSENDOZ ON 11/19/14 AT 5:12 PM

Uber is tracking its drivers in China, will fire anyone attending taxi protests
by David DClicks - Jul 14, 2016, 10:46 PM

Uber’s ‘Hell’ program tracked and targeted Lyft drivers
The top-secret program apparently ran for two years.
by Monika Matyi, businessinsider.com | 04/12/17 in Transportation | 41 Comments 2936 Shares
ORide: Goals

• Privacy for riders and drivers
• Usability and accountability
Adversarial Assumptions

Service Provider (SP)

Active

Riders (R)

Internet

Honest-but-curious

Drivers (D)

Active
Attacks Considered

• Riders and drivers might assault each other

• SP might perform large-scale privacy-sensitive inferences

• SP might perform targeted attacks
Desired Usability and Accountability Features

• Riders and drivers are accountable for their behaviors
  - The SP can identify misbehaving users
  - Even in extreme cases: e.g., kidnapping

• Usability:
  - Easy payment
  - Reputation ratings for riders and drivers
  - Retrieval of lost items
ORide Overview

Rider (anonymous)

Drivers (anonymous)

Notations:
ACs: Anonymous credentials
cert: Digital certificate

cert_R
ACs

Ride set-up

Oblivious ride matching

Accountability mechanisms

Ride termination

Payment

Reputation rating
Oblivious Ride Matching - Intuition

- **Goal:** Rider can select the closest driver without revealing their locations to the SP
Oblivious Ride Matching

**Goal:** Rider can select the closest driver without revealing their locations to the SP

**Rider** (anonymous)

- Certificate $\text{cert}_R$
- Attributes $\text{ACs}$

**SP**

1. Zone $D$

**Drivers** (anonymous)

- Certificate $\text{cert}_D$
- Attributes $\text{ACs}$

**Process:***

1. Rider generates $(k_p, k_s)$
2. Zone $R$, $E_{k_p}\{\text{loc}_R\}$, $k_p$
3. $k_p$
4. $E_{k_p}\{\text{loc}_{D_i}\}$
5. $E_{k_p}\{\Delta_{R-D_1}\}$, ..., $E_{k_p}\{\Delta_{R-D_n}\}$
6. Index $\text{ind}$
7. Notify driver with index $\text{ind}$

**Computation:**

$E_{k_p}\{\Delta_{R-D_i}\}$

where $\Delta_{R-D_i} = [\text{dist}([\text{loc}_R, \text{loc}_{D_i}])]^2$
SP Computes $E_{k_p} \{ \Delta_{R-D_i} \}$

with $\Delta_{R-D_i} = [\text{dist}(\text{loc}_R, \text{loc}_{D_i})]^2 = (x_R - x_{D_i})^2 + (y_R - y_{D_i})^2$

• Use Somewhat-Homomorphic Encryption (SHE) scheme by Fan and Vercauteren

$$\begin{cases} R_t = Z_t[X]/(X^d + 1) & \text{Plaintexts} \\ R_q = Z_q[X]/(X^d + 1) & \text{Ciphertexts and keys} \end{cases}$$

• How?

\[
\begin{array}{cccccccc}
 x_R & 0 & \cdots & 0 & \cdots & 0 \\
 - \\
 x_{D_i} & 0 & \cdots & 0 & \cdots & 0 \\
\end{array}
\]

\[d \text{ slots (each in } Z_t)\]

\[
\begin{array}{cccccccc}
 x_R - x_{D_i} & 0 & \cdots & 0 & \cdots & 0 \\
 \times \\
 x_R - x_{D_i} & 0 & \cdots & 0 & \cdots & 0 \\
\end{array}
\]

\[
\begin{array}{cccccccc}
 (x_R - x_{D_i})^2 & 0 & \cdots & 0 & \cdots & 0 \\
 + \\
 (y_R - y_{D_i})^2 & 0 & \cdots & 0 & \cdots & 0 \\
\end{array}
\]

\[
\begin{array}{cccccccc}
 \Delta_{R-D_i} & 0 & \cdots & 0 & \cdots & 0 \\
\end{array}
\]

![Naive](repeat_n_times) Repeat $n$ times ($n = \text{No. of drivers}$)  
→ Inefficient and overhead
SP Computes $E_{k_p}\{\Delta_{R-D_i}\}$

with $\Delta_{R-D_i} = [\text{dist}(\text{loc}_R, \text{loc}_{D_i})]^2 = (x_R - x_{D_i})^2 + (y_R - y_{D_i})^2$

• **Optimization Goal:** 1 computation/request for R and SP

<table>
<thead>
<tr>
<th>Naïve</th>
<th>Optimized</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta_{R-D_0}$</td>
<td>$\begin{bmatrix} \Delta_{R-D_0} &amp; \ldots &amp; \Delta_{R-D_{n-1}} &amp; 0 &amp; \ldots &amp; 0 \end{bmatrix}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta_{R-D_i}$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>$\Delta_{R-D_i}$</td>
<td>$0$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td></td>
</tr>
<tr>
<td>$\Delta_{R-D_{n-1}}$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

- Ciphertext packing: to pack drivers’ ciphertexts into 1 ciphertext
- Number-Theoretic Transform (NTT): to support coefficient-wise multiplication
Oblivious Ride Matching

- Rider R selected driver D without revealing their locations to the SP

- How R and D share their locations w/ each other?
- Accountability?
  - R and D might assault each other
  - R might not want to pay for the ride
  - Dispute on the fare
Accountability Mechanisms

• **Goal:** Rider and driver are accountable for their behaviors

Rider (anonymous)

- $\text{cert}_R$
- ACs

Driver (anonymous)

- $\text{cert}_D$
- ACs

**Keep locally or share with trusted peers**

(8) Secure channel: exchange $\text{rep}_R$, $\text{rep}_D$, $\text{loc}_R$, $\text{loc}_D$, $\text{cert}_R$ and $\text{cert}_D$, gen. a PIN

(9) $\text{loc}_D$ in real time

(10) Proximity check & validation of secure channel

(11) Driver’s info: plate number and profile photo

(12) Fare report: $\text{sig}_{R-D}\{\text{day, fare, cert}_R, \text{cert}_D\}$

The ride begins

--- Secure channel
--- Proximity channel

In proximity of $\text{loc}_R$
Payment and Reputation Rating

- **Goal**: The two operations do not break the anonymity of the ride

---

1. Rider
   - \(\text{cert}_R\)
   - \(\text{ACs}\)

2. Driver
   - \(\text{cert}_D\)
   - \(\text{ACs}\)

---

13. Fare report: \(\text{sig}_{R-D}\{\text{day, fare, cert}_R, \text{cert}_D\}\)

14. Charge R from his credit card

15a. Rate D

15b. Rate R

---

End of the day
Protocol Analysis

• Information observed by the SP

<table>
<thead>
<tr>
<th>Ride DB</th>
<th>Payment DB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone of the pick-up</td>
<td>Rider’s ID</td>
</tr>
<tr>
<td>Obfuscated pick-up time</td>
<td>Fare</td>
</tr>
</tbody>
</table>

• Targeted attacks by the SP

Pick-up loc. and time of a specific rider → Drop-off location

Drop-off locations and times are never reported to the SP
Protocol Analysis

• Information observed by the SP

<table>
<thead>
<tr>
<th>Zone of the pick-up</th>
<th>Obfuscated pick-up time</th>
</tr>
</thead>
</table>

Ride DB

<table>
<thead>
<tr>
<th>Rider’s ID</th>
<th>Driver’s ID</th>
<th>Fare</th>
<th>Day of the ride</th>
</tr>
</thead>
</table>

Payment DB

• Large-scale inferences by the SP

Home/work addresses → Sensitive activities

**Lower-bound** anonymity set = No. of rides from the same zone on the same day
Evaluation

• Data-set: taxi rides in NYC in Oct. 2013 (15 millions rides)

• How practical and efficient is ORide?
  - Per-ride overhead
  - Riders’ anonymity vs. drivers’ bandwidth
  - Effect of Euclidean distance on ride-matching optimality

Census Tracts (CT) and Neighborhood Tabulation Areas (NTA)
Implementation

• Feature the oblivious ride matching algorithm
• SHE parameters: > 112 bits of security

- Ciphertexts and keys

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(124 bits)</td>
<td>(124 bits)</td>
<td>...</td>
<td>(124 bits)</td>
<td>...</td>
<td>(124 bits)</td>
</tr>
</tbody>
</table>

- Plaintexts

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(20 bits)</td>
<td>(20 bits)</td>
<td>...</td>
<td>(20 bits)</td>
<td>...</td>
<td>(20 bits)</td>
</tr>
</tbody>
</table>

- In C++ using NFLlib
- No SSE or AVX optimizations

4096 slots
Per-Ride Overhead

- Measured on Intel i5-4200U, 2.6 GHz, 6 GB RAM
- Bandwidth overhead

<table>
<thead>
<tr>
<th>Setting</th>
<th>Rider</th>
<th>Driver</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Upload (KB)</td>
<td>Download (KB)</td>
</tr>
<tr>
<td>S1</td>
<td>372</td>
<td>761856</td>
</tr>
<tr>
<td>S2</td>
<td>372</td>
<td>186</td>
</tr>
</tbody>
</table>

- Computation time

<table>
<thead>
<tr>
<th>Setting</th>
<th>Rider</th>
<th>Driver</th>
<th>SP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gen. keys (ms)</td>
<td>Encrypt (ms)</td>
<td>Decrypt (ms)</td>
</tr>
<tr>
<td>S1</td>
<td>1.51</td>
<td>2.6</td>
<td>7823.4</td>
</tr>
<tr>
<td>S2</td>
<td>1.51</td>
<td>2.6</td>
<td>2.2</td>
</tr>
</tbody>
</table>

The optimized approach significantly reduces bandwidth and computation overhead
Riders’ Anonymity Set vs. Drivers’ Bandwidth

• Zones:
  - Queens + Bronx
  - Brooklyn + Staten Island
  - Manhattan CTs
  - Manhattan NTAs

Large anonymity set and reasonable bandwidth requirements
Conclusion

• ORide: practical and privacy-preserving
  - Strong privacy guarantees
  - Negligible overhead

• Still offers key RHS features:
  - Accountability
  - Easy payment
  - Reputation scores

http://oride.epfl.ch

Source: http://blog.cyberghostvpn.com/avoid-being-tracked-online/