Thwarting Longitudinal Location Exposure Attacks in Advertising Ecosystem via Edge Computing

Le Yu\textsuperscript{1†}, Shufan Zhang\textsuperscript{2†}, Lu Zhou\textsuperscript{1}
Yan Meng\textsuperscript{1}, Suguo Du\textsuperscript{1}, Haojin Zhu\textsuperscript{1}

1 Shanghai Jiao Tong University
2 University of Waterloo
Outline

• Background
  • Motivation
  • System
  • Evaluation
Background

- Location-based Advertising (LBA)
  - Growing market (12.8% expected annual growth)
    - Finer-grained, personalized service
    - High return-on-investment (RoI) rate
Background

• Location-based Advertising (LBA)
  • Growing market (12.8% expected annual growth)
    • Finer-grained, personalized service
    • High return-on-investment (RoI) rate
  • Business model

The business model and data flow of LBA
Background

• Location-based Advertising (LBA)
  • Growing market (12.8% expected annual growth)
    • Finer-grained, personalized service
    • High return-on-investment (RoI) rate
  • Business model
  • Types of location targeting
    • Countries targeting
    • Areas targeting
    • Radius targeting (finest-grained)

<table>
<thead>
<tr>
<th>Companies</th>
<th>Minimal Radius</th>
<th>Maximal Radius</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>5 km</td>
<td>65 km</td>
</tr>
<tr>
<td>Microsoft</td>
<td>1 mile / 1 km</td>
<td>800 miles / 800 km</td>
</tr>
<tr>
<td>Facebook</td>
<td>1 mile</td>
<td>50 miles</td>
</tr>
<tr>
<td>Tencent</td>
<td>500 m</td>
<td>25 km</td>
</tr>
</tbody>
</table>
Background

• Location-based Advertising (LBA)
  • Growing market (12.8% expected annual growth)
    • Finer-grained, personalized service
    • High return-on-investment (ROI) rate
  • Business model
  • Types of location targeting
    • Countries targeting
    • Areas targeting
    • Radius targeting (finest-grained)
  • Privacy becomes prominent issue

<table>
<thead>
<tr>
<th>Companies</th>
<th>Minimal Radius</th>
<th>Maximal Radius</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>5 km</td>
<td>65 km</td>
</tr>
<tr>
<td>Microsoft</td>
<td>1 mile / 1 km</td>
<td>800 miles / 800 km</td>
</tr>
<tr>
<td>Facebook</td>
<td>1 mile</td>
<td>50 miles</td>
</tr>
<tr>
<td>Tencent</td>
<td>500 m</td>
<td>25 km</td>
</tr>
</tbody>
</table>
Motivating Example

• People have stable mobility pattern
  • Location entropy
  • We can recover user’s mobility pattern

A user's 7-day mobility pattern

Entropy = \sum_{i=1}^{M} \frac{f_i}{\text{sum}} \log \frac{\text{sum}}{f_i}

Calls for location obfuscation mechanism

88.8% of users' location entropy is less than 2
Related Work

• Location Privacy
  • Privacy protection with theoretical guarantee
  • Differential Privacy (DP) [DMNS06]

Related Work

- Location Privacy
  - Privacy protection with theoretical guarantee
  - Differential Privacy (DP) [DMNS06]
    - Location trajectory synthesis (e.g., DPT [HCMP15])
    - Location obfuscation (e.g., Geo-IND [ABCP13])


Outline

• Background
• Motivation
• System
• Evaluation
Motivation

• Huge gap between theoretical Geo-IND and real-world privacy issues in LBA!
• *New attack*: Longitudinal location exposure attack
Motivation

One-time obfuscation mechanism:
• Planar Laplace mechanism / Geo-Indistinguishability [ABCP13]

Longitudinal location exposure attack

Longitudinal Attack in LBA

Set-up:
- Raw location check-ins
Longitudinal Attack in LBA

Set-up:
- Raw location check-ins
- Obfuscate the location check-ins using planar Laplace mechanism
Longitudinal Attack in LBA

Set-up:
- Raw location check-ins
- Obfuscate the location check-ins using planar Laplace mechanism

Recover Top Location:
- **Step 1 Clustering**: Cluster locations check-ins based on connectivity (distance threshold)
Longitudinal Attack in LBA

Set-up:
- Raw location check-ins
- Obfuscate the location check-ins using planar Laplace mechanism

Recover Top Location:
- **Step 1 Clustering:** Cluster locations check-ins based on connectivity (distance threshold)
- **Step 2 Trimming:** drop out locations whose distance is larger than cluster radius

Cluster radius $r_\alpha$, $\Pr[dist(p, q) > r_\alpha] \leq \alpha$
Outline

• Background
• Motivation
• System
• Evaluation
Insight

Permanent obfuscation

- Insight: users are refrained to their top locations
- Challenge: how to reduce utility loss

\( AOI \): area of interest  
\( AOR \): area of request

Utilization rate  
\[
UR = \frac{AOI \cap AOR}{AOI}
\]

Advertiser efficacy  
\[
AE = \Pr [ad \in AOI | ad \in AOR]
\]
Insight

Permanent obfuscation

- Insight: users are refrained to their top locations
- Challenge: how to reduce utility loss
- Multiple obfuscated locations

\( AOI \): area of interest
\( AOR \): area of request

Utilization rate \( UR = \frac{AOI \cap AOR}{AOI} \)

Advertiser efficacy \( AE = Pr[ad \in AOI|ad \in AOR] \)
Privacy Definition

Generalize geo-IND to $(r, n, \varepsilon, \delta)$-geo-IND

**Mapping** $p \rightarrow \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix}$

$r$-Neighboring

For all pair of locations $p_0, p_1$, we say $p_0, p_1$ are $r$-neighboring if the Euclidean distance between $p_0$ and $p_1$ is less than $r$, that is $\text{dist} (p_0, p_1) < r$.

$$
Pr \left[ p_0 \rightarrow \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix} \right] \leq e^\varepsilon Pr \left[ p_1 \rightarrow \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix} \right] + \delta
$$

$(r, n, \varepsilon, \delta)$-geo-IND
System Overview

- How to avoid repeated obfuscation in long-term usage
- How to provide tighter error bound
- How to provide reasonable utility
System Design

Module 1 Location Management

- User check-ins are not directly used for LBA
- Passively collect users' location data
- Compute top frequent locations
System Design

Module 1  Location Management

Module 2  Location Obfuscation

User Check-ins

$\eta$-Frequent Location Set

Processing

Clustering

$\eta$-Frequent Location Set

Processing

$n$-fold Gaussian Mechanism

Location Mapping Table

Stores the top locations and their obfuscated locations
System Design

Module 1: Location Management
- User Check-ins
- $\eta$-Frequent Location Set
- Processing
  - Clustering
  - $\eta$-Frequent Location Set

Module 2: Location Obfuscation
- Location Mapping Table
- Processing
  - $n$-fold Gaussian Mechanism

Module 3: Output Selection
- Location to report
- Processing
  - Resampling
  - Post-processing without privacy loss
n-fold Gaussian Mechanism

• $n$ independent Gaussian random variables $N(p, \sigma^2)$

$$(q_1, \ldots, q_n) = (p + X_1, \ldots, p + X_n)$$

Challenge: solving $\sigma$ to satisfy $(r, n, \epsilon, \delta)$-geo-IND

- Naïve composition: $\epsilon' = \frac{\epsilon}{n}, \delta' = \frac{\delta}{n}$

$$\sigma = \frac{nr}{\epsilon} \sqrt{\ln \frac{1}{(n\delta)^2} + \frac{\epsilon}{n}}$$
n-fold Gaussian Mechanism

• $n$ independent Gaussian random variables $N(p, \sigma^2)$

$$(q_1, \ldots q_n) = (p + X_1, \ldots p + X_n)$$

Challenge: solving $\sigma$ to satisfy $(r, n, \varepsilon, \delta)$-geo-IND

- Naïve composition: $\varepsilon' = \frac{\varepsilon}{n}$, $\delta' = \frac{\delta}{n}$

$$\sigma = \frac{nr}{\varepsilon} \sqrt{\ln \left( \frac{1}{(n\delta)^2} + \frac{\varepsilon}{n} \right)}$$

- Sufficient Statistics
The following statements are equivalent:
  • Releasing $(q_1, \ldots q_n)$ satisfies $(r, n, \varepsilon, \delta)$-geo-IND
  • Releasing the sufficient statistic of $(q_1, \ldots q_n)$ satisfies $(r, 1, \varepsilon, \delta)$-geo-IND

$$\sigma = \frac{\sqrt{nr}}{\varepsilon} \sqrt{\ln \left( \frac{1}{\delta^2} + \varepsilon \right)}$$

Tighter error bound!
Outline

• Background
• Motivation
• System
• Evaluation
Evaluation

Dataset

• We collect 37,262 mobile users in Shanghai from June 1, 2019 to May 31, 2021
• The size ranges from 20 to 11,435 check-ins per user.
• The dataset are from a real-world RTB transaction-log dataset

Parameter settings.

• $\delta = 0.01$ and $\varepsilon \in \{1, 1.5\}$
• The indistinguishable radius $r = 500$ m, 600 m, 700 m, 800 m.
• The targeting radius we choose is $R = 5$ km
What's the Attack success rate in one-time obfuscation and permanent obfuscation?

**Observation 1**
Attack success rate of one-time obfuscation (200 m):
top-1 locations: 75% for $l = \ln 2$, 90% for $l = \ln 4$ and $\ln 6$, top-2 locations: more than 50% for $l = \ln 4$ and $\ln 6$. 
What's the performance of the n-fold Gaussian mechanism?

Observation 2
The n-fold Gaussian mechanism outperforms the naïve post-processing mechanism and the plain DP composition-based Gaussian mechanism.

Parameters: $r = 500, \varepsilon = 1, \delta = 0.01$
What's the impact of the obfuscation number $n$ and privacy parameters?

Observation 3
The utilization rate increase with $n$

Parameters: $\epsilon = 1$ or $1.5$, $\delta = 0.01$
What's the efficacy of Edge-PrivLocAd?

**Observation 4**
The efficacy do not significantly decrease with $n$

**Parameters**: $\epsilon = 1, \delta = 0.01$
Scalability of Edge-PrivLocAd

• Emulation with Raspberry Pi 3

<table>
<thead>
<tr>
<th>Number of Users</th>
<th>2000</th>
<th>4000</th>
<th>8000</th>
<th>16000</th>
<th>32000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Time (s)</td>
<td>340</td>
<td>627</td>
<td>1166</td>
<td>2090</td>
<td>4014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Users</th>
<th>2000</th>
<th>4000</th>
<th>8000</th>
<th>16000</th>
<th>32000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Time (ms)</td>
<td>90</td>
<td>175</td>
<td>350</td>
<td>698</td>
<td>1377</td>
</tr>
</tbody>
</table>

The emulation shows our system is scalable in edge environment
The processing time for obfuscation and output selection is reasonable
Takeaways

• **New Attack.** Existing geo-IND mechanisms cannot be directly applied to long-term location exposure settings, e.g., LBA.

• **New Mechanism.** The $n$-fold Gaussian mechanism is proposed to achieve tight composition bound (optimized utility) when releasing $n$ locations simultaneously.

• **New System.** Edge-PrivLocAd is built to provide long-term location privacy management for LBA.

• Extensive experiments have shown the effectiveness and the efficiency of the proposed system.