DYNAMIC DIFFERENTIAL LOCATION PRIVACY WITH PERSONALIZED ERROR BOUNDS

LEI YU, LING LIU AND CALTON PU
COLLEGE OF COMPUTING
GEORGIA INSTITUTE OF TECHNOLOGY
Location based services and Privacy issues

Location based services
- Marketing
  - FANDANGO
  - yelp
  - Google places
  - FOURSQUARE
  - Groupon
- Social Networking
  - Facebook
  - twitter
  - tinder
  - loopt
- Gaming
  - GPS wan
  - Pokemon GO
- Sports, Navigation, Health, Media, etc.

New Tinder Security Flaw Exposed Users’ Exact Locations for Months

Man Accused of Stalking Ex-Girlfriend With GPS

Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset
Location Privacy Protection

Anonymization

- $K$-anonymity
- trusted third-party anonymization server
Location Privacy Protection

Location Obfuscation

- Use a fake location instead of the true location
- User-centric
- Client-side

\[ p(x' | x) = \Pr(\text{pseudo\_location}=x' | \text{actual\_location}=x) \]
Location Obfuscation

Privacy Notion

Randomized mechanism

Utility

\[ p(x^\prime \mid x) = \Pr(\text{pseudo\_location}=x^\prime \mid \text{actual\_location}=x) \]
Existing Techniques

- Privacy Notions:
  - Expected inference error
  - Geo-indistinguishability
• The expected distance between the user’s real location and the location guessed by the adversary.

Prior distribution $\pi(x)$ of the user being at location $x$.

Given observation $x'$, the probability of actual location being $x$

$$\Pr(x|\pi(x')) = \frac{\pi(x)f(x'|x)}{\sum_{x' \in \mathcal{X}} \pi(x)f(x'|x)}$$
Geo-indistinguishability

For any two points $x, y$ in the protection circular area of radius $r$ centered at the actual location, by $\epsilon g = \epsilon / 2r$

$$\frac{f(x' | x)}{f(x' | y)} \leq e^{\epsilon}$$
### Existing Techniques

- **Privacy Notions:**

<table>
<thead>
<tr>
<th>Expected inference error</th>
<th>Geo-indistinguishability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian inference</td>
<td>differential privacy</td>
</tr>
<tr>
<td>Rely on a specific prior distribution of user’s real location</td>
<td>only depends on the mechanism and does not depend on any prior</td>
</tr>
<tr>
<td>Not robust against any other prior distribution</td>
<td>Adding noise regardless of any prior can be inefficient and insufficient for privacy protection</td>
</tr>
</tbody>
</table>
Our work

• Limitation of Geo-indistinguishability
• Two-phase location obfuscation framework
  – Adaptive noise level for different locations with guaranteeing a minimum level of inference error
  – Customizability
    • Instantly specify his privacy preference for his current location
    • Existing mechanisms are computed statically once for all, and cannot efficiently support customizability
Experimental Illustration

- Existing mechanisms
  - Optimal Bayesian mechanism [R. Shokri et al., 2012]
  - Optimal geo-indistinguishable mechanism [N. E. Bordenabe et al., 2014]
Experimental Illustration

50 regions with prior probability >0

Dataset: GeoLife GPS Trajectories dataset  
Formatted as in [N. E. Bordenabe et al., 2014]

Two mechanisms that achieve the same location privacy in terms of overall expected inference error weighted by prior probability
- Geo-indistinguishability

Planar Laplacian Mechanism, $\Pr(\text{pseudo-location in blue circle}) \geq 95\%$

Not Adaptable: Uniform noise level either insufficient location protection at some skewed locations in terms of prior information or excessive noise for protection at other locations.
Two-phase framework

- Combine expected inference error and Geo-indistinguishability

Diagram:

- Prior distribution
  - $\pi$

- Searching protection region
  - $x$
  - $E_{\downarrow m}$

- Protection location set
  - $\Phi$

- Exponential mechanism
  - $\epsilon$
  - $x'$

- True location
  - $x$

- Minimum Inference error
  - $E_{\downarrow m}$
Relationship between two privacy notions

• Geo-indistinguishability
  – Any two locations \( x, y \) in the protection region \( \Phi \),

\[
\frac{f(x|x)}{f(y|y)} \leq e^{\epsilon}
\]

• Lower bound of conditional expected inference error

\[
\min_{x} \sum_{x \in \Phi} \Pr(x) d(x,x) \geq e^{-\epsilon} \min_{x} \sum_{x \in \Phi} \pi(x) \frac{d(x,x)}{\sum_{y \in \Phi} \pi(y)}
\]
• **Theorem:** For a location obfuscation mechanism that achieves \( \epsilon \)-differential privacy on protection location set \( \Phi \), if \( E(\Phi) \geq e^{-\epsilon} E\downarrow m \), the optimal inference attack using any observed pseudo-location \( x' \), the expected inference error \( \geq E\downarrow m \).

\[
E(\phi) = \min_{\pi} \sum_{x \in \Phi} \sum_{y \in \Phi} d(x, y) \pi(x) / \pi(y)
\]
Phase I: Search Protection Region

- $E(\Phi) \geq e^{\epsilon} E \downarrow m$
- Hilbert-curve based searching
  - Larger diameter of protection location set indicates higher noise level
  - Improvement with multiple rotated Hilbert curves
Phase II: Exponential mechanism

• Given the user’s location $x$ and location protection set $\Phi$, the exponential mechanism selects and outputs a pseudo-location $x'$ with probability proportional to $\exp(-\epsilon d(x, x')/2D)$, where $D$ is the diameter of $\Phi$. 
Evaluation

• Comparison with existing mechanisms on location privacy

EM - Laplacian-like mechanism uniform noise level

(a) Optimal inference attack
\[ \hat{x} = \arg \min_{\hat{x} \in X} \sum_{x \in X} \Pr(x|x') d_p(\hat{x}, x) \]

(b) Bayesian inference attack
\[ \hat{x} = \arg \max_{x \in X} \Pr(x|x') \]
Evaluation

• Comparison with joint mechanism on location privacy
Quality loss: the average distance between the fake location and the real location.
• Geo-indistinguishability + prior information
• Adaptively adjust noise level of different privacy according to prior distribution
• Customizability
Thank you!

Q&A
Expected inference error

Conditional expected inference error
\[ \sum_{x, x' \in \mathcal{X}} \pi(x) f(x, x') \]
the distance between the estimation and the actual location

\[ h(x | x') \] - Probability of guessing \( x \) as the user’s actual location, given that \( x' \) is observed

Optimal inference attack: \( x = \arg\min_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} \pi(x) f(x, x') \)

Bayesian inference attack: \( x = \arg\max_{x \in \mathcal{X}} \Pr(x | x') \)

Unconditional expected inference error
\[ \sum_{x, x' \in \mathcal{X}} \pi(x) f(x, x') \]

Quality loss
\[ \sum_{x, x' \in \mathcal{X}} \pi(x) f(x, x') \]