PGLP: Customizable and Rigorous Location Privacy through Policy Graph

Yang Cao¹, Yonghui Xiao², Shun Takagi¹, Li Xiong², Masatoshi Yoshikawa¹, Yilin Shen³, Jinfei Liu², Hongxia Jin³, and Xiaofeng Xu²

¹ Kyoto University, ² Emory University, ³ Samsung Research America
Outline

• Motivation
  — why we need a customizable and rigorous location privacy model.

• Our Solution: Policy Graph based Location Privacy (PGLP)
  — a flexible interface for location privacy to tune privacy-utility tradeoffs.

• PGLP for Location Trace Release
  — challenges and countermeasures when using PGLP continuously.

• Experiments

• Conclusion & Future work
Outline

• Motivation
  — why we need a customizable and rigorous location privacy model.

• Our Solution: Policy Graph based Location Privacy (PGLP)
  — a flexible interface for location privacy to tune privacy-utility tradeoffs.

• PGLP for Location Trace Release
  — challenges and countermeasures when using PGLP continuously.

• Experiments

• Conclusion & Future work
Motivation

Location data: **valuable but sensitive**

- **Useful** in our daily life for Location-based Service (LBS)
- **Fundamental** in research areas IoT, Crowdsourcing, Smart City..
Motivation

Location data: **valuable but sensitive**

• risky for an individual
  ▸ reveal many sensitive info: identification, home, office, lifestyle, hobbies...

• risky for companies who utilize user location data

Google Map “frequently visited locations”

Europe’s huge privacy fines against Marriott and British Airways are a warning for Google and Facebook

General Data Protection Regulation
Motivation

How to Protect Location Privacy

- **general idea**: add **uncertainty** to the true location

Location Privacy Model

Privacy Definition: metrics of privacy

LPPM: achieve a privacy metrics with high utility

“what is weather tomorrow near my location?”

- **My location = Kyoto University** (better utility)
- **My location = Kyoto City** (better privacy)

**general research goal**: better tradeoff between **privacy** and **utility**
Motivation

Existing Location Privacy Definitions

- Extended from K-anonymity
  - Location k-anonymity, [MobiSys03].
  - Mix zone, [PerCom03].
  - The New Casper [VLDB06].
  - Maximum arrival boundary [TKDE12].

- Extended from Differential Privacy (DP)
  - Geo-Indistinguishability [CCS13].
  - $\delta$-location set privacy [CCS15].
Motivation

Existing Location Privacy Definitions are Not Sufficient

- K-anonymity based Location privacy is not rigorous
- L-diversity argue K-anonymity has flaws
- T-closeness say: L-diversity has flaws
- ....
- Existing DP-based location privacy is not customizable
  - Only use one parameter $\epsilon$ to control the privacy-utility trade-off.
  - However, different LBS may have different requirements on privacy or utility.
Motivation
Different LBS, Different Utility Requirement

• City-level weather forecast
  • Query: which city is the user in?
  • High utility when the noisy location is in the same city of the true location.

• Location-based advertising
  • Query: what kind of loc. (shopping mall/restaurant) is the user in?
  • High utility when the noisy location has the same category of the true location.

• Location-based Social Network
  • Query: where is my nearest friend?
  • High utility when the distance between two noisy locations is similar to the distance between the true locations.
Outline

• Motivation
  — why we need a customizable and rigorous location privacy model.

• Our Solution: Policy Graph based Location Privacy (PGLP)
  — a flexible interface for location privacy to tune privacy-utility tradeoffs.

• PGLP for Location Trace Release
  — challenges and countermeasures when using PGLP continuously.

• Experiments

• Conclusion & Future work
Our Solution
Intuitions for a Customizable and Rigorous Location Privacy

• Inspired by Blowfish Privacy [SIGMOD14], the privacy-utility Tradeoff can be fine-tuned by “Privacy Policies”:
  • secrets: what are the secrets that we need to protect?
  • constraints: what does the adversary know?

• However, Blowfish Privacy cannot be directly applied in location privacy.
  • Location privacy: single user, point query on single record (location).
  • Statistical privacy: multiple users, aggregate query on a database.

• How to formalize Location Privacy Policy and how to achieve it?
Our Solution: PGLP
Location Policy Graphs

- How to formalize these policies? — Location Privacy Policy Graph
  
  - **Nodes**: the user’s possible locations.
  
  - **Edges**: the two connected locations need to be indistinguishable to the adversary.

- The right policy is a good fit for “Location-based advertising”
  
  - High utility if the noisy location has the same category of the true location.

  - Policy: “allowing the app to access the semantic category (e.g., a restaurant or a shop) of a user’s location but ensuring indistinguishability among locations with the same category.”
Our Solution: PGLP

Definition

- **key idea**: only satisfy the indistinguishability defined in the given policy graph.

- Location Policy Graph:

  **Definition 3 (Location Policy Graph).** A location policy graph is an undirected graph $G = (S, E)$ where $S$ denotes all the locations (nodes) and $E$ represents indistinguishability (edges) between these locations.

- Policy Graph-based Location Privacy:

  **Definition 6 ($\{\epsilon, G\}$-Location Privacy).** A randomized algorithm $A$ satisfies $\{\epsilon, G\}$-location privacy iff for all $z \subseteq \text{Range}(A)$ and for all pairs of neighbors $s$ and $s'$ in $G$, we have $\frac{\Pr(A(s) = z)}{\Pr(A(s') = z)} \leq e^\epsilon$. 
Our Solution: PGLP

Definition

- PGLP is a generalization of DP-based location privacy definitions.
  - it reduces to Geo-Indistinguishability [CCS13] and δ-location set privacy [CCS15] under different configuration of the policy graph.

Theorem 1. An algorithm satisfying \( \{\epsilon, \mathcal{G}_1\}\)-location privacy also achieves \( \epsilon \)-Geo-Indistinguishability.

Theorem 2. An algorithm satisfying \( \{\epsilon, \mathcal{G}_2\}\)-location privacy also achieves \( \delta \)-Location Set privacy.

![Fig. 2: Two examples of location policy graphs.](image)
Our Solution: PGLP

Mechanisms

- **key idea:** calibrate the sensitivity w.r.t. a given policy graph.

**Algorithm 1** Policy-based Laplace Mechanism (P-LM)

**Require:** $\epsilon$, $\mathcal{G}$, the user’s true location $s$.
1. Calculate $S_f^G = \sup || (f(s) - f(s'))/d_G(s, s') ||_1$ for all $s' \in \mathcal{N}^\infty(s)$;
2. Perturb location $z' = f(s) + [\text{Lap}(S_f^G/\epsilon), \text{Lap}(S_f^G/\epsilon)]^T$;
3. **return** a location $z \in S$ that is closest to $z'$ on the map.

**Algorithm 2** Policy-based Planar Isotropic Mechanism (P-PIM)

**Require:** $\epsilon$, $\mathcal{G}$, the user’s true location $s$.
1. Calculate $K(\mathcal{G}) = \text{Conv} || (f(s) - f(s'))/d_G(s, s') ||_1$ for all $s' \in \mathcal{N}^\infty(s)$;
2. $z' = f(s) + Y$ where $Y$ is two-dimension noise drawn by Eq.(1) with sensitivity hull $K(\mathcal{G})$;
3. **return** a location $z \in S$ that is closest to $z'$ on the map.
Outline

• Motivation
  — why we need a customizable and rigorous location privacy model.

• Our Solution: Policy Graph based Location Privacy (PGLP)
  — a flexible interface for location privacy to tune privacy-utility tradeoffs.

• PGLP for Location Trace Release
  — challenges and countermeasures when using PGLP continuously.

• Experiments

• Conclusion & Future work
PGLP for Continuous Release

Challenges

- The user possible location set may change over time.

- **Location Exposure** under constrained domain:
  If the user is at s5, the attacker may be able to figure out her true loc.

- Not all of the disconnected node will lead to Location Exposure, which also depends on the mechanism.
PGLP for Continuous Release

Countermeasure: Risk Detection and Policy Repair algorithm

- Detect Isolated Node in a policy graph
  - **Isolated Node**: the disconnected node that causes location exposure.
- Repair a policy graph with high utility.
  - **Key idea**: add an edge to protect the isolated node.

**Algorithm 3 Finding Isolated Node**

```
Require: \( G, C \), disconnected node \( s_i \in C \).
1: \( \Delta f^G = \bigvee_{s_j, s_k \in \mathbb{X}_C} (f(s_j) - f(s_k)) \);
2: \( K(G^C) \leftarrow \text{Conv}(\Delta f^G) \);
3: for all \( s_j \in C, s_j \neq s_i \) do
4:   if \( \text{Conv}(\Delta f^G, f(s_j) - f(s_i)) = K(G^C) \) then
5:     return false
6:   end if
7: end for
8: return true
```

See our paper for more details.
An end-to-end Location Trace release framework

- Pipelines the *calculation of constrained domains, isolated node detection, policy graph repair, and private location release mechanism*.

- Utilizing HMM model (assume transition and initial probabilities are known)

Fig. 5: Private location trace release via HMM.
Outline

• Motivation
  — why we need a customizable and rigorous location privacy model.

• Our Solution: Policy Graph based Location Privacy (PGLP)
  — a flexible interface for location privacy to tune privacy-utility tradeoffs.

• PGLP for Location Trace Release
  — challenges and countermeasures when using PGLP continuously.

• Experiments

• Conclusion & Future work
Experiments

• How different location policy graphs affect the privacy-utility tradeoffs?

• Settings:

  • Two types of location policy graphs:
    • “block-graph”: $G_{k9}, G_{k16}, G_{k25}$ suitable for weather apps
    • “category-graph”: $G_{poi}$ suitable for location-based advertising

  • Three types of Utilities
    • $E_{eu}$: Euclidean distance between noisy and true locations.
    • $E_r$: L0 distance between range queries on noisy and true locations, like “whether the released location is in the same region with the true location” suitable for weather apps
    • $E_{poi}$: L0 distance between category queries on noisy and true locations, like “whether the released location is the same category with the true location”. suitable for location-based advertising
Experiments

Verified that we can flexibly design suitable policy w.r.t. the desired utility & privacy.

- Observations: $G_{k9}$ is best for $E_{eu}$ and $E_r$; $G_{poi}$ is the best for $E_{poi}$.

Fig. 7: Utility of different policy graphs.

(check the paper for more experimental results)
Outline

• Motivation
  — why we need a customizable and rigorous location privacy model.

• Our Solution: Policy Graph based Location Privacy (PGLP)
  — a flexible interface for location privacy to tune privacy-utility tradeoffs.

• PGLP for Location Trace Release
  — challenges and countermeasures when using PGLP continuously.

• Experiments

• Conclusion & Future work
Conclusion & Future Work

• Takeaway
  • PGLP provides a rich interface for privacy-utility tradeoff in location privacy.

• Future directions
  • Design advanced mechanisms for PGLP
  • Design optimal policy graphs for location-based applications, such as spatial crowdsourcing.