Thwarting Longitudinal Location Exposure Attacks in Advertising Ecosystem via Edge Computing

Le Yu[1†], Shufan Zhang[2†], Lu Zhou[1]
Yan Meng[1], Suguo Du[1], Haojin Zhu[1]

1 Shanghai Jiao Tong University

2 University of Waterloo

Outline

- Background
- Motivation
- System
- Evaluation

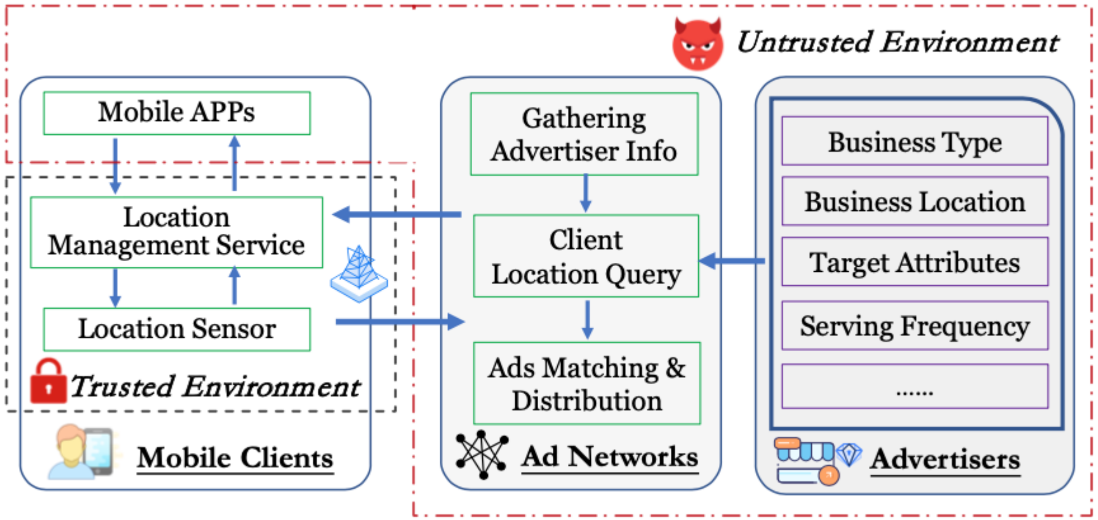
Background

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

Positioning Technologies

Background

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

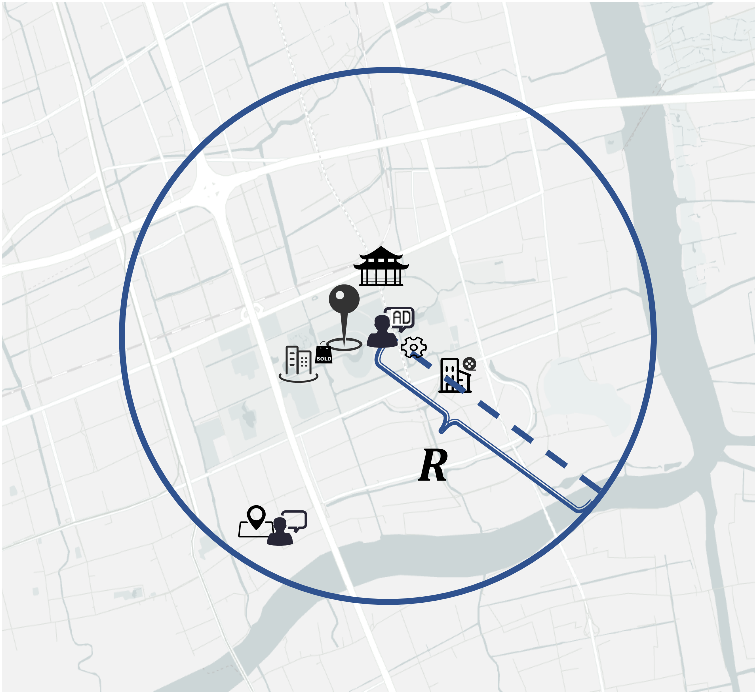


The business model and data flow of LBA

Positioning Technologies

Background

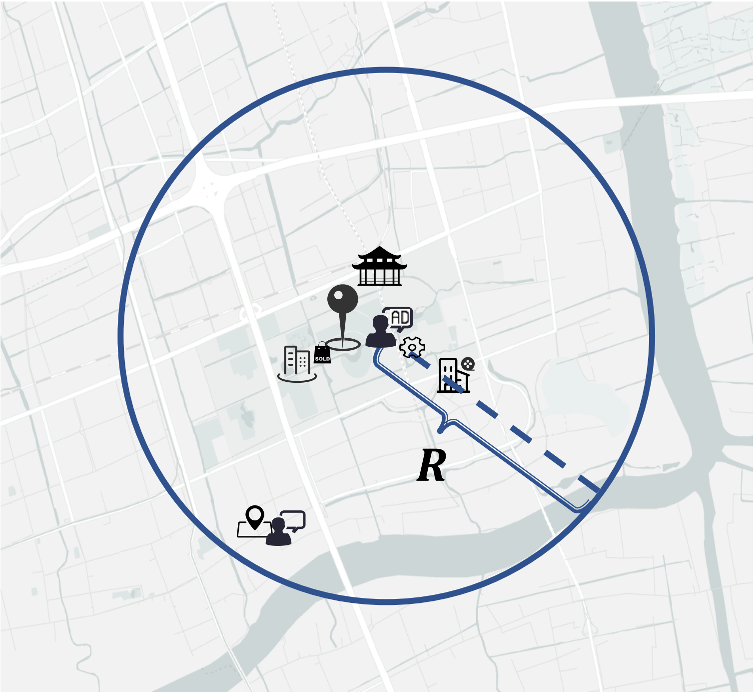
- Location-based Advertising (LBA)
  - Growing market (12.8% expected annual growth)
    - Finer-grained, personalized service
    - High return-on-investment (RoI) rate
  - Business model
  - Types of location targeting
    - Countries targeting
    - Areas targeting
    - Radius targeting (finest-grained)



Companies	Minimal Radius	Maximal Radius
Google	5 km	65 km
Microsoft	1 mile / 1 km	800 miles / 800 km
Facebook	1 mile	50 miles
Tencent	500 m	25 km

Background

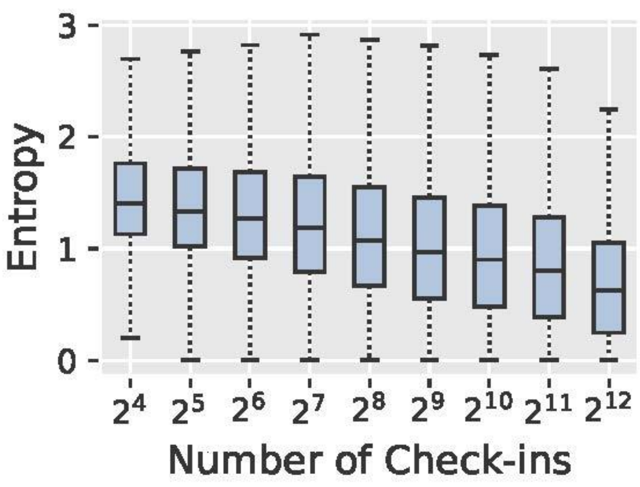
- Location-based Advertising (LBA)
  - Growing market (12.8% expected annual growth)
    - Finer-grained, personalized service
    - High return-on-investment (RoI) rate
  - Business model
  - Types of location targeting
    - Countries targeting
    - Areas targeting
    - Radius targeting (finest-grained)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



Companies	Minimal Radius	Maximal Radius
Google	5 km	65 km
Microsoft	1 mile / 1 km	800 miles / 800 km
Facebook	1 mile	50 miles
Tencent	500 m	25 km

Motivating Example

- People have stable mobility pattern
  - Location entropy
  - We can recover user's mobility pattern



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

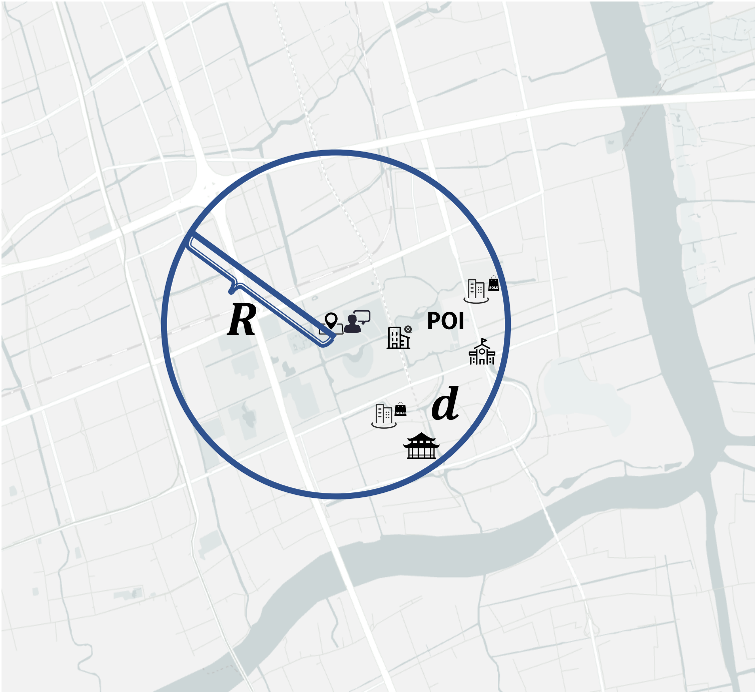
88.8% of users' location entropy is less than 2

A user's 7-day mobility pattern

Calls for location obfuscation mechanism

Related Work

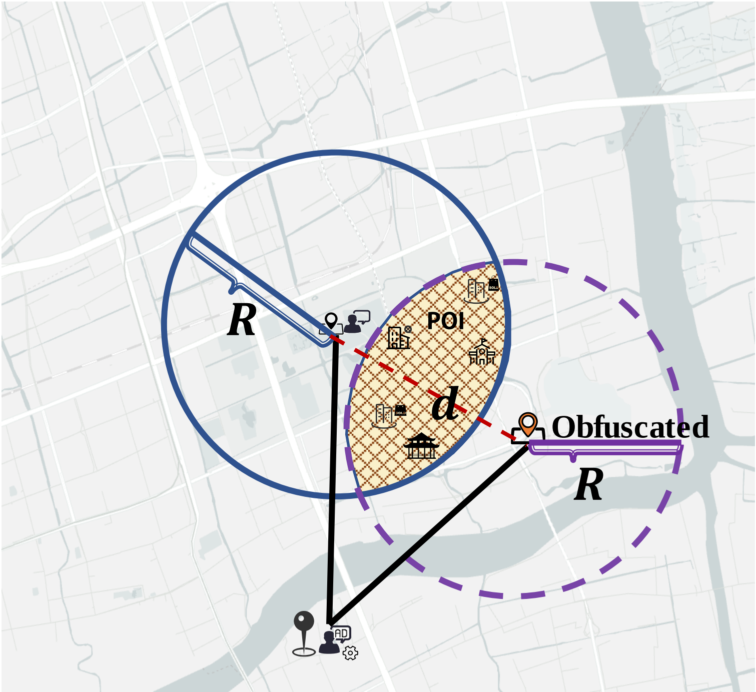
- Location Privacy
  - Privacy protection with theoretical guarantee
  - Differential Privacy (DP) [DMNS06]



Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006, March). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference* (pp. 265-284). Springer, Berlin, Heidelberg.

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])



Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006, March). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference* (pp. 265-284). Springer, Berlin, Heidelberg.
He, X., Cormode, G., Machanavajjhala, A., Procopiuc, C. M., & Srivastava, D. (2015). DPT: differentially private trajectory synthesis using hierarchical reference systems. *Proceedings of the VLDB Endowment*, *8*(11), 1154-1165.
Andrés, M. E., Bordenabe, N. E., Chatzikokolakis, K., & Palamidessi, C. (2013, November). Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security* (pp. 901-914).

Outline

- Background
- Motivation
- System
- Evaluation

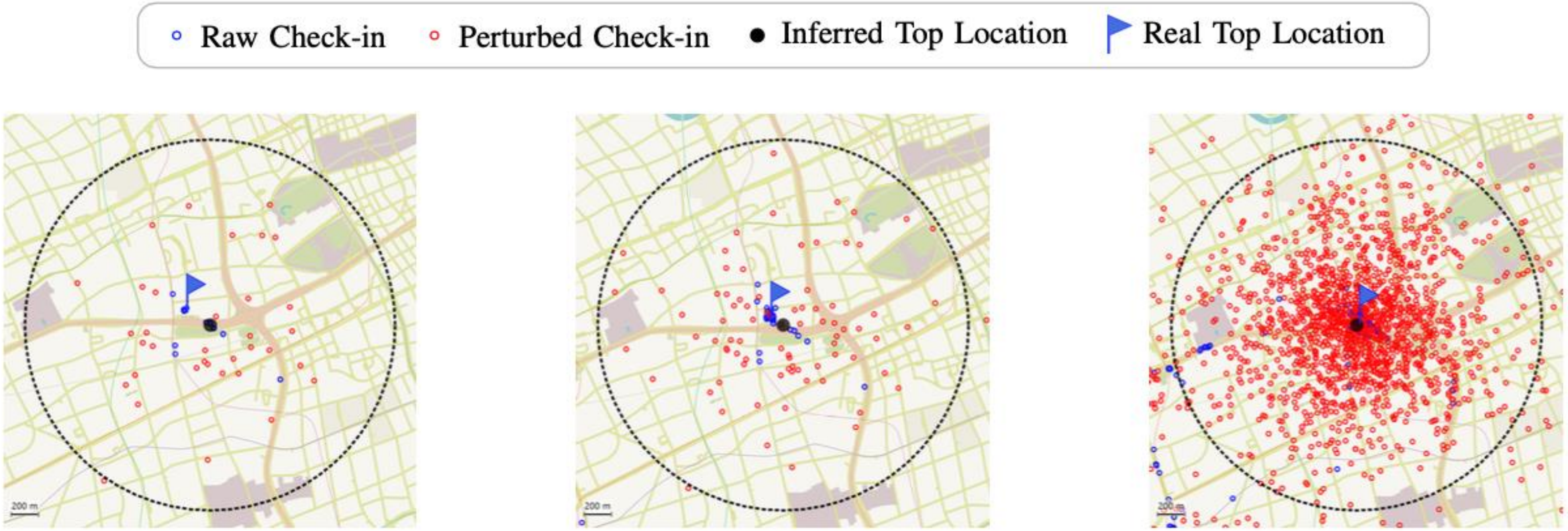
Motivation

- 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]

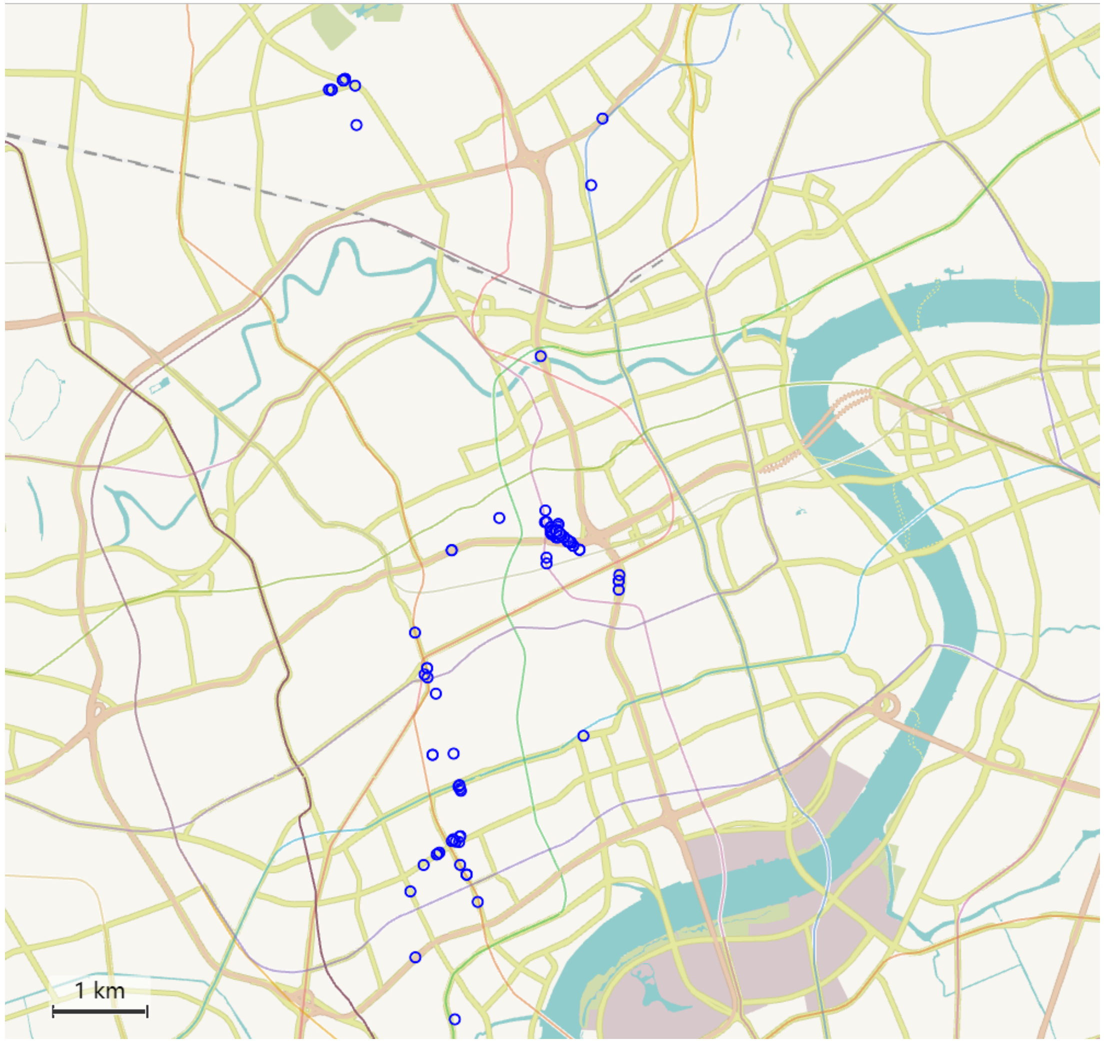


(a) One-week Data (b) One-month Data (c) Full Year Data

Longitudinal location exposure attack

Andrés, M. E., Bordenabe, N. E., Chatzikokolakis, K., & Palamidessi, C. (2013, November). Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security* (pp. 901-914).

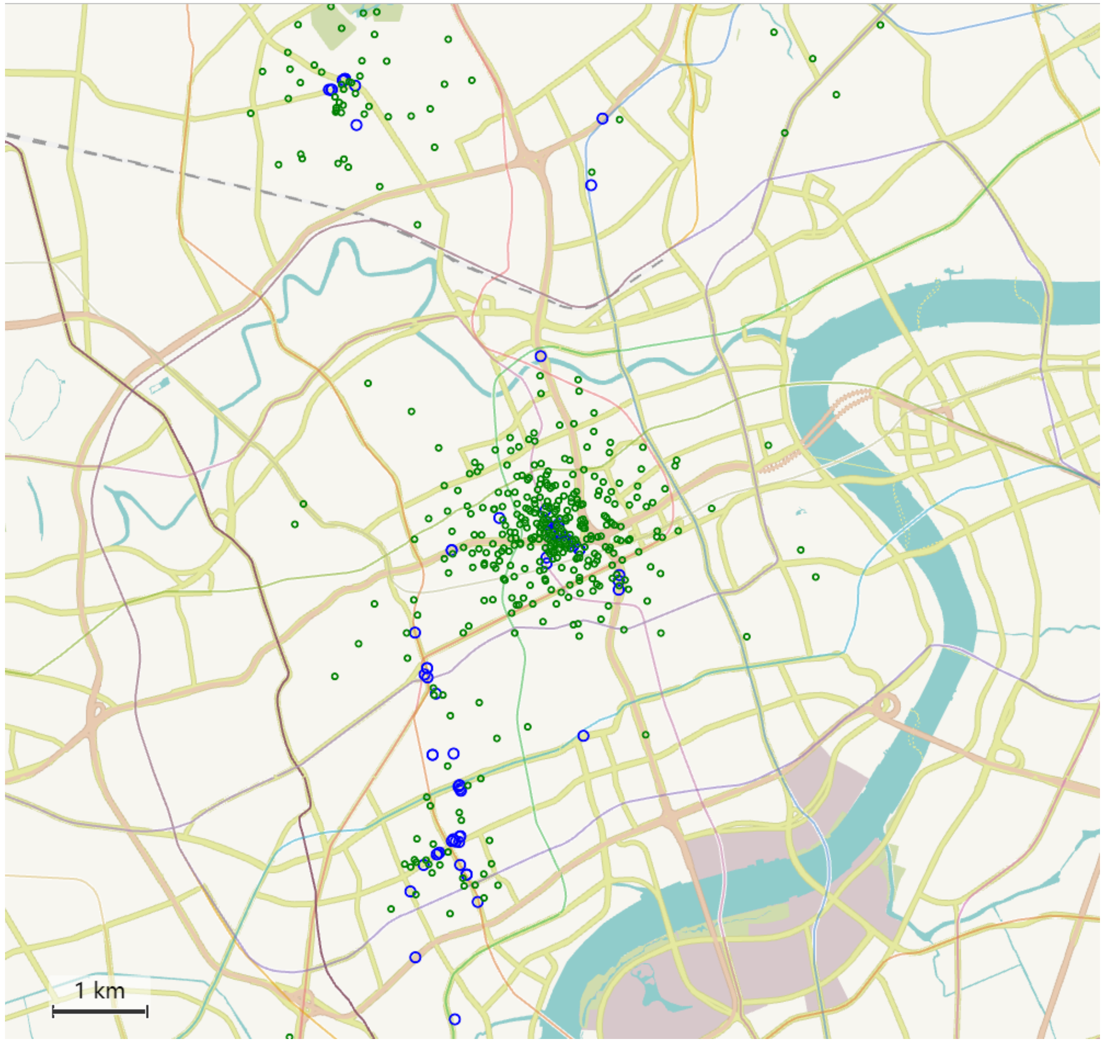
Longitudinal Attack in LBA



Set-up:

- Raw location check-ins

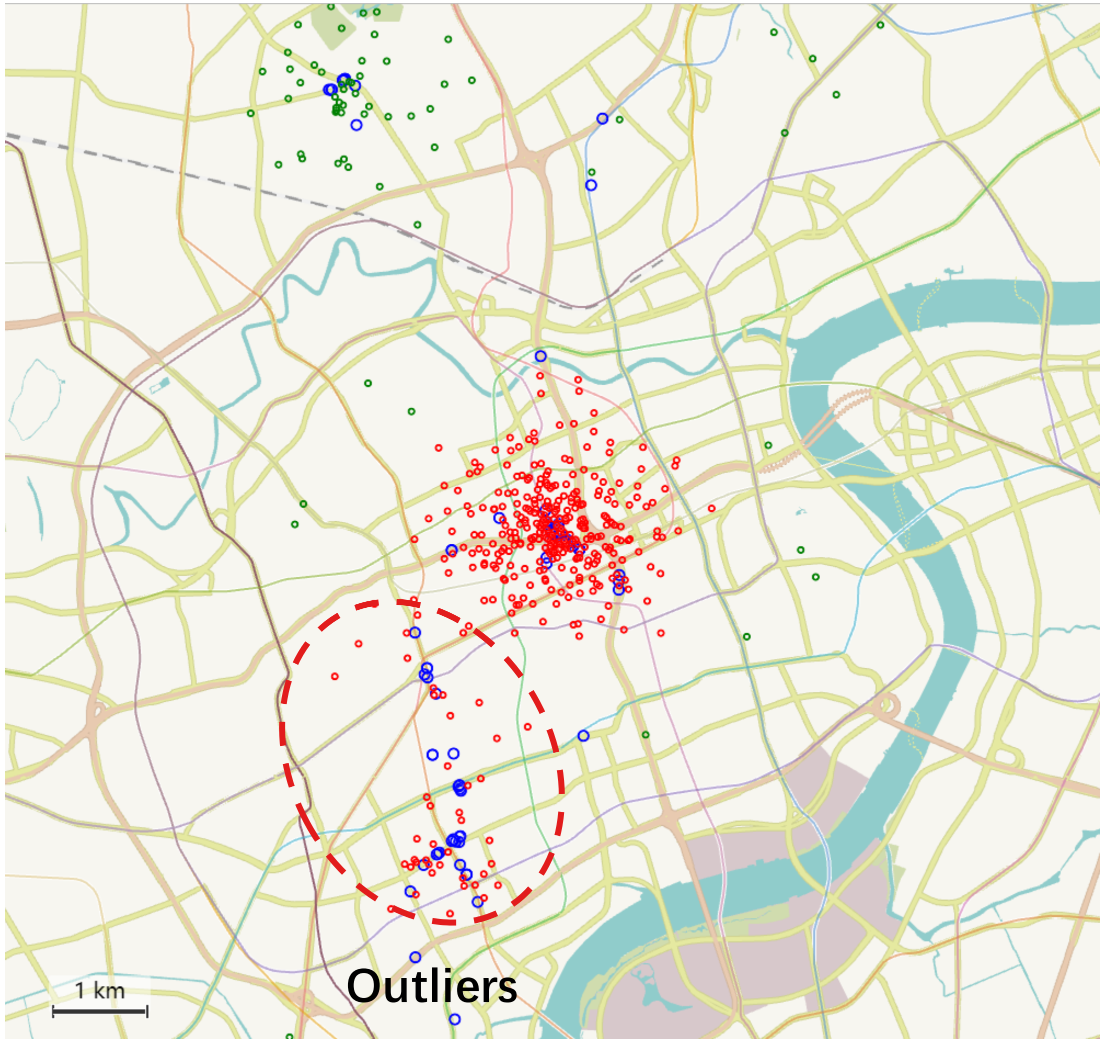
Longitudinal Attack in LBA



Set-up:

- Raw location check-ins
- Obfuscate the location check-ins using planar Laplace mechanism

Longitudinal Attack in LBA



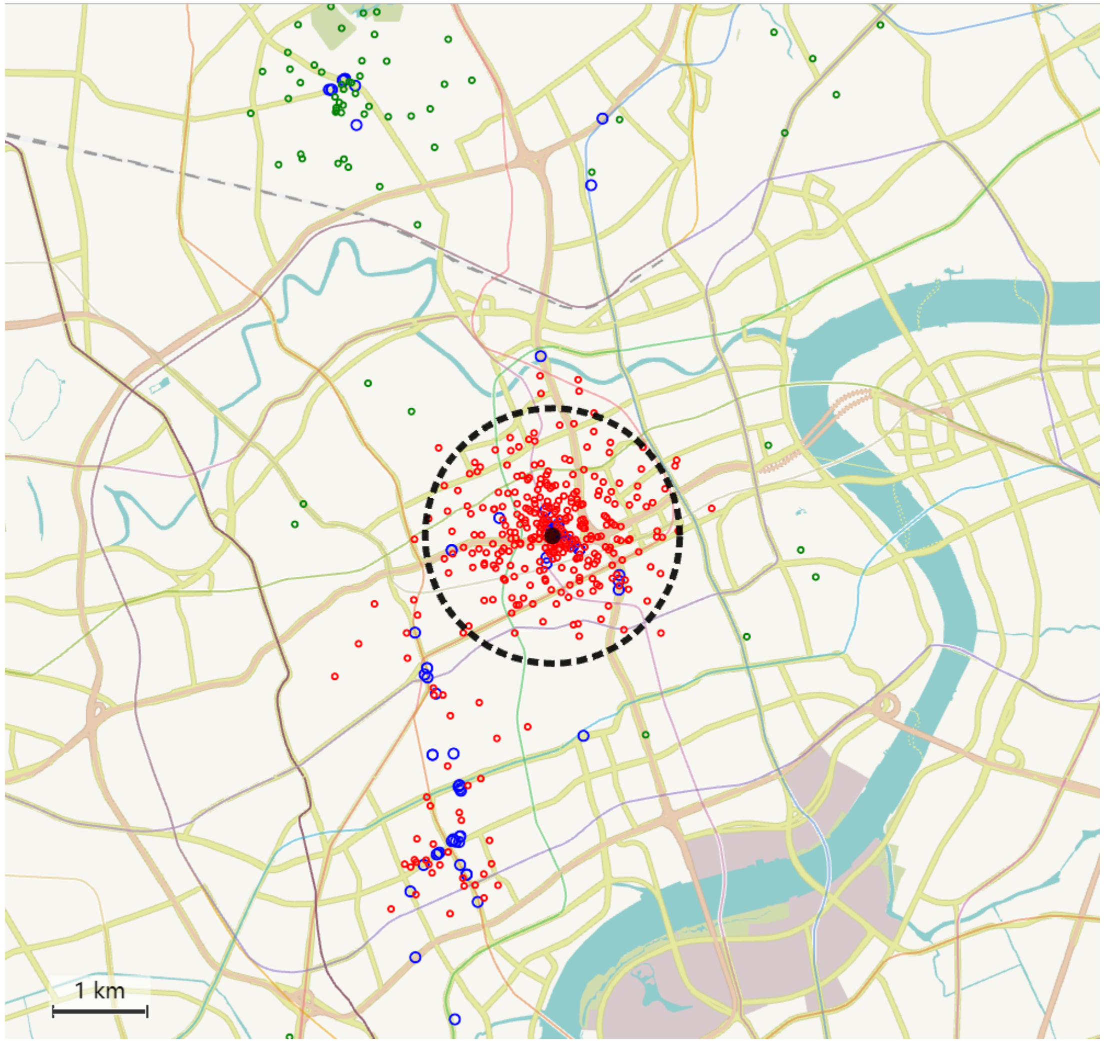
Set-up:

- 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_α, $\Pr[dist(p,q) > r_\alpha] \le \alpha$

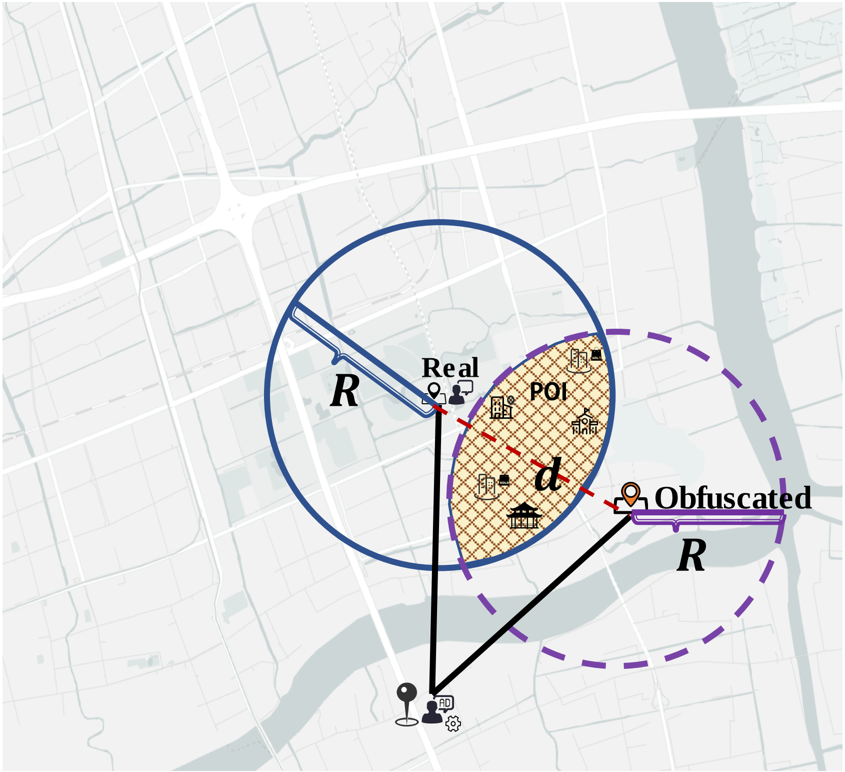
Outline

- 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

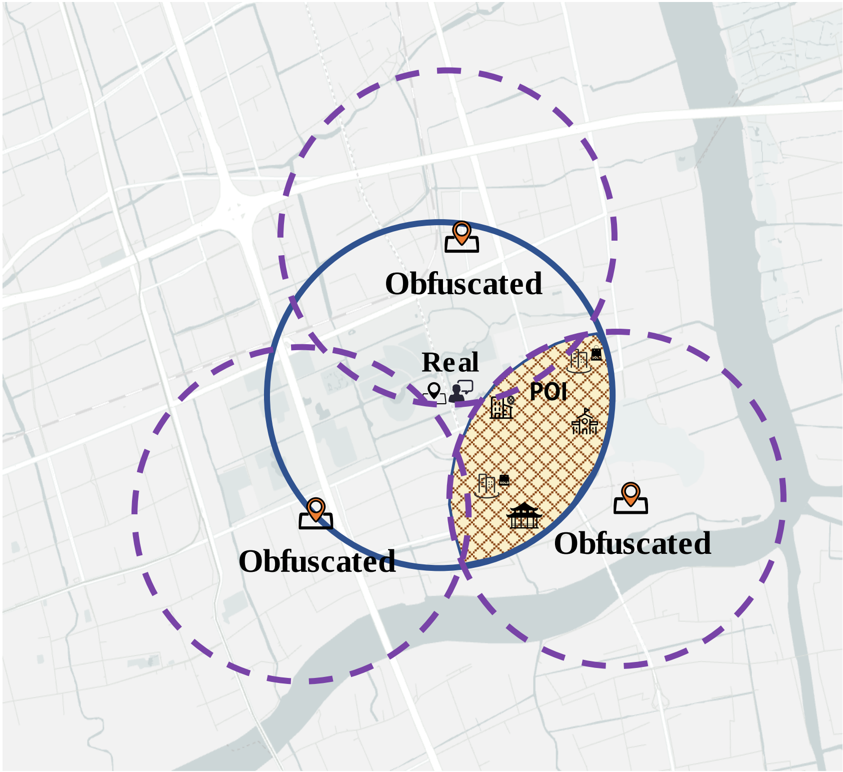
Utilization rate $UR = \frac{AOI \cap AOR}{AOI}$

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



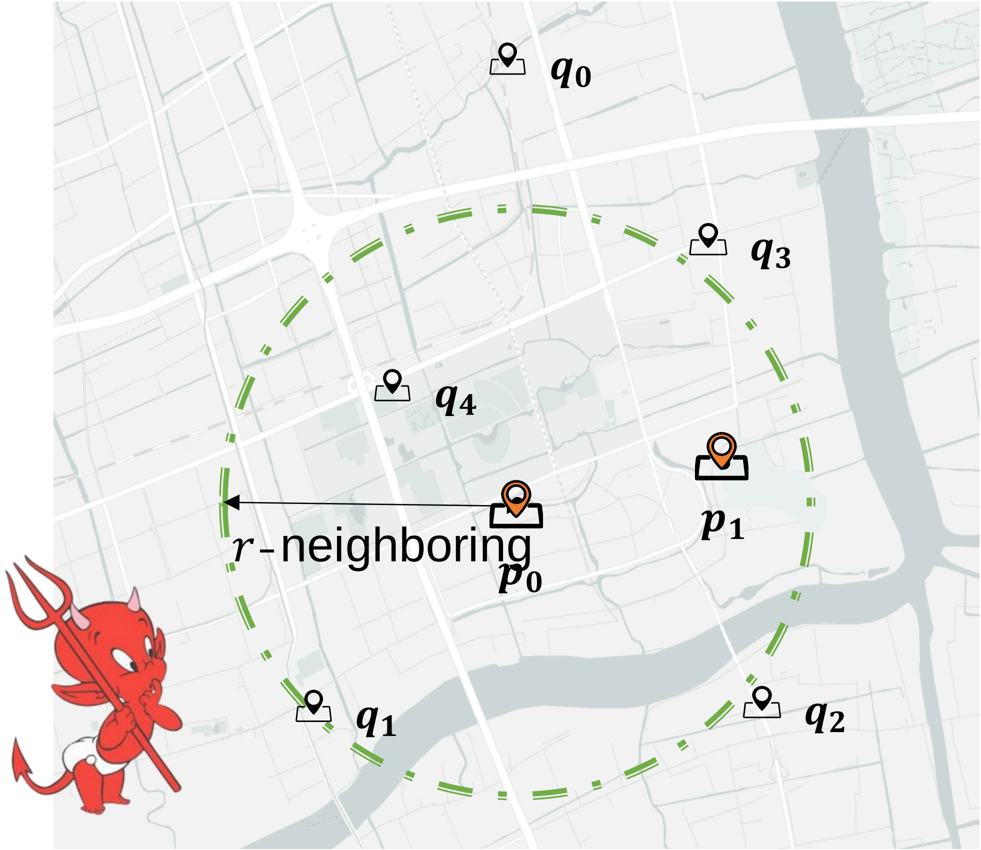
AOI : area of interest
AOR : area of request
Utilization rate $UR = \frac{AOI \cap AOR}{AOI}$
Advertiser efficacy $AE = \Pr[ad \in AOI | ad \in AOR]$

Privacy Definition

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

Mapping $\boldsymbol{p} \rightarrow \begin{pmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \\ \vdots \\ \boldsymbol{q}_n \end{pmatrix}$

r-Neighboring
For all pair of locations $\boldsymbol{p}_0$, $\boldsymbol{p}_1$, we say $\boldsymbol{p}_0$, $\boldsymbol{p}_1$ are r-neighboring if the Euclidean distance between $\boldsymbol{p}_0$ and $\boldsymbol{p}_1$ is less than r, that is $dist(\boldsymbol{p}_0, \boldsymbol{p}_1) < r$.



$$\boldsymbol{Pr}\left[\boldsymbol{p}_0 \rightarrow \begin{pmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \\ \vdots \\ \boldsymbol{q}_n \end{pmatrix}\right] \leq \boldsymbol{e}^{\boldsymbol{\varepsilon}} \boldsymbol{Pr}\left[\boldsymbol{p}_1 \rightarrow \begin{pmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \\ \vdots \\ \boldsymbol{q}_n \end{pmatrix}\right] + \boldsymbol{\delta}$$

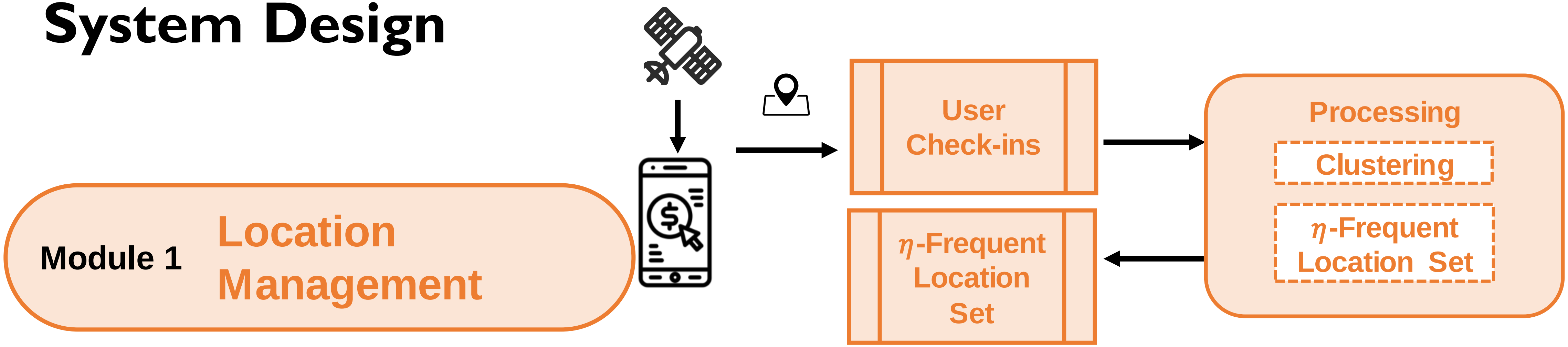
$(r, n, \varepsilon, \delta)$-geo-IND

System Overview

Edge-PrivLocAd

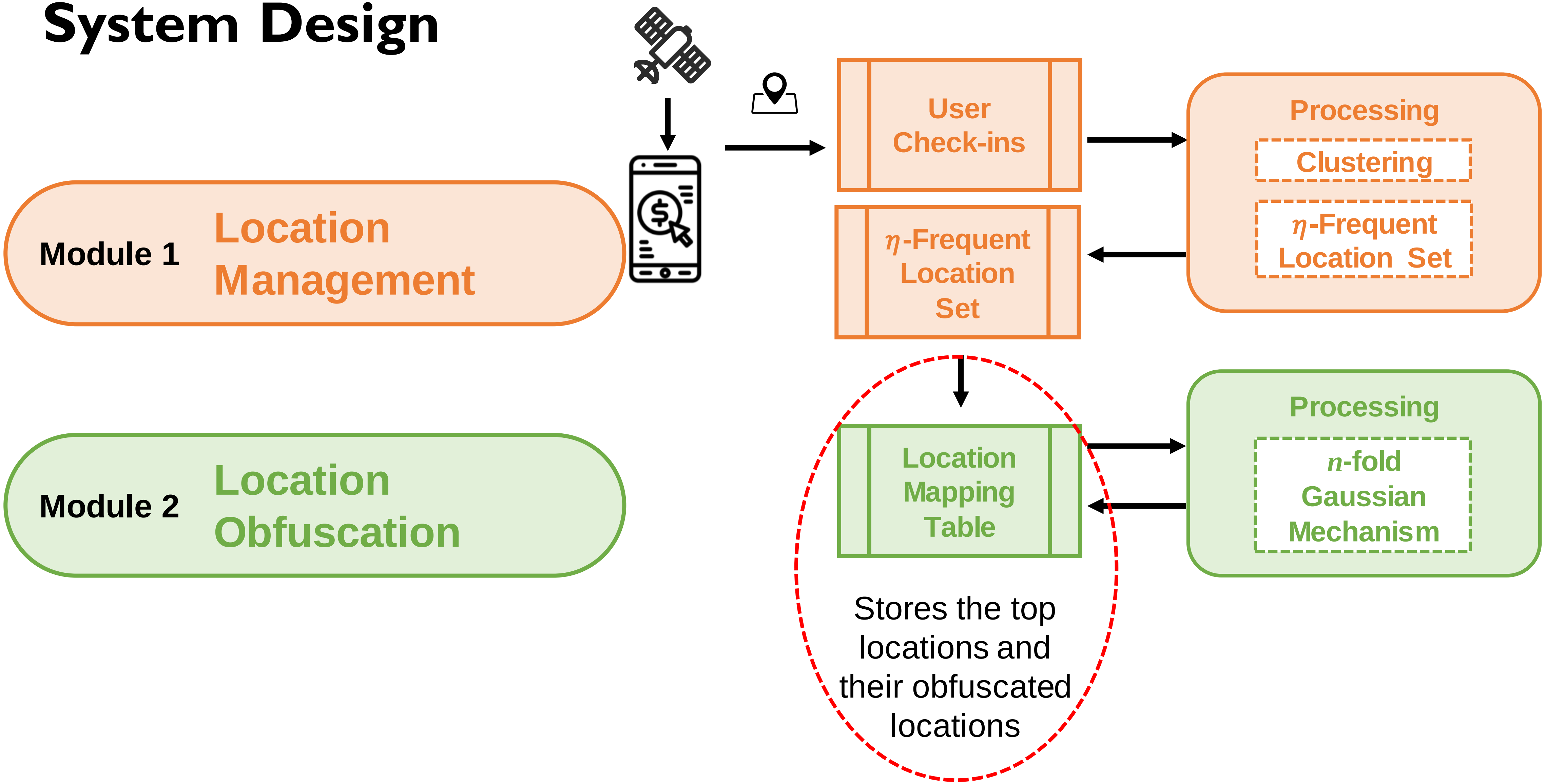
- How to avoid repeated obfuscation in long-term usage — **Module 1**: **Location Management**
- How to provide tighter error bound — **Module 2**: **Location Obfuscation**
- How to provide reasonable utility — **Module 3**: **Output Selection**

System Design

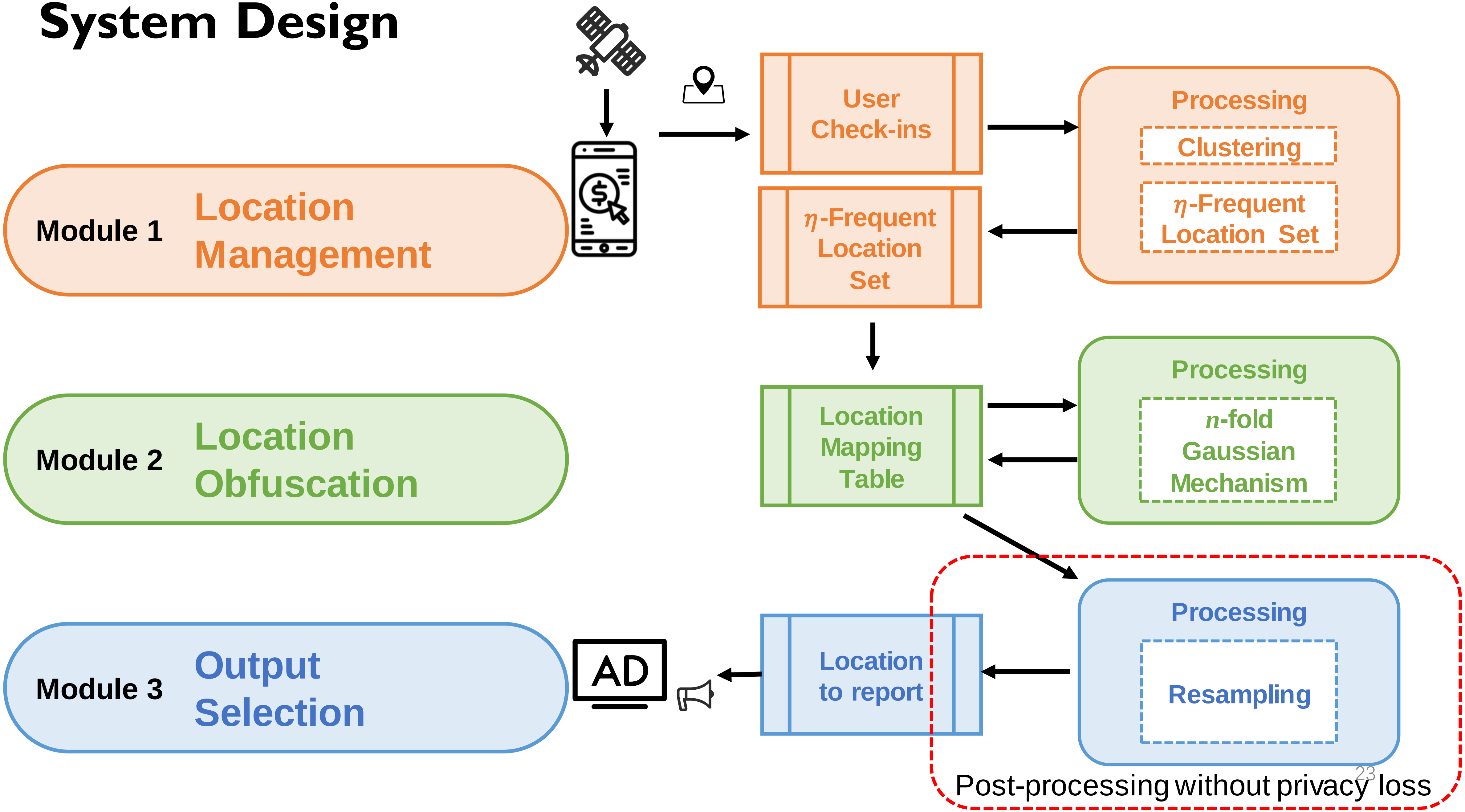


- User check-ins are **not directly** used for LBA
- Passively collect users' location data
- Compute top frequent locations

System Design



System Design



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

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

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

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

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

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

- ☐ 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

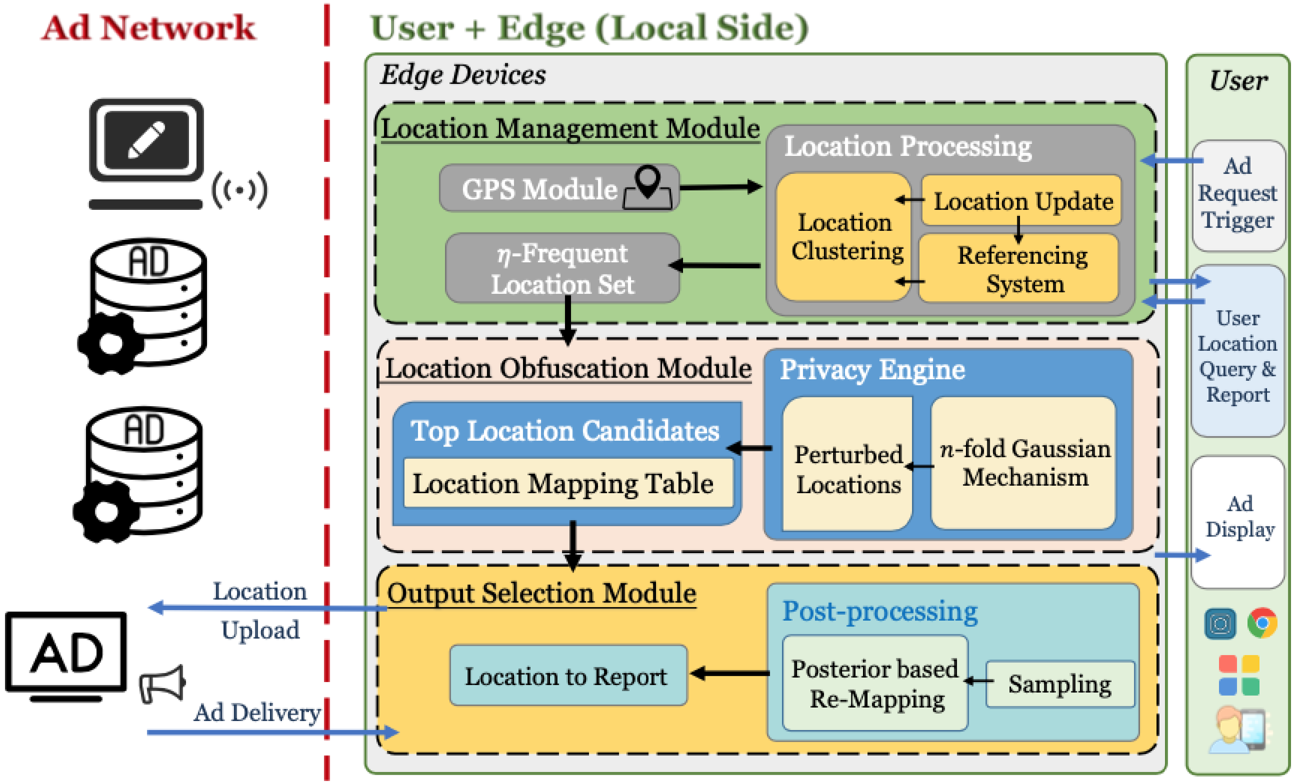
$$\sigma = \frac{\sqrt{n}r}{\varepsilon}\sqrt{\ln\frac{1}{\delta^2} + \varepsilon}$$

Tighter error bound!

Outline

- Background
- Motivation
- System
- Evaluation

Evaluation



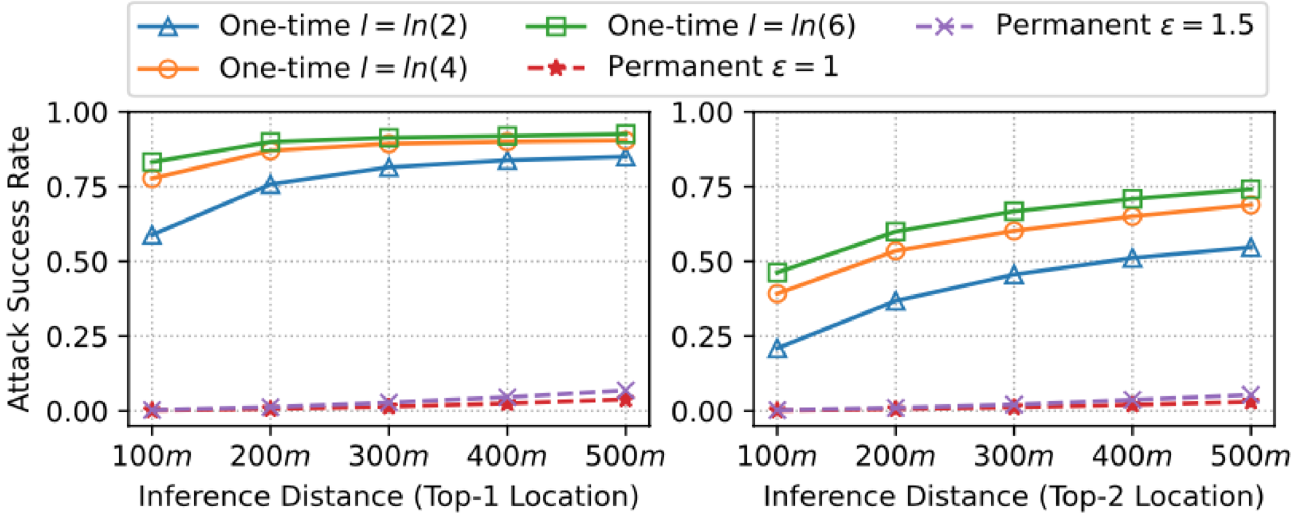
Dataset

- We collect 37,262 mobiles 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

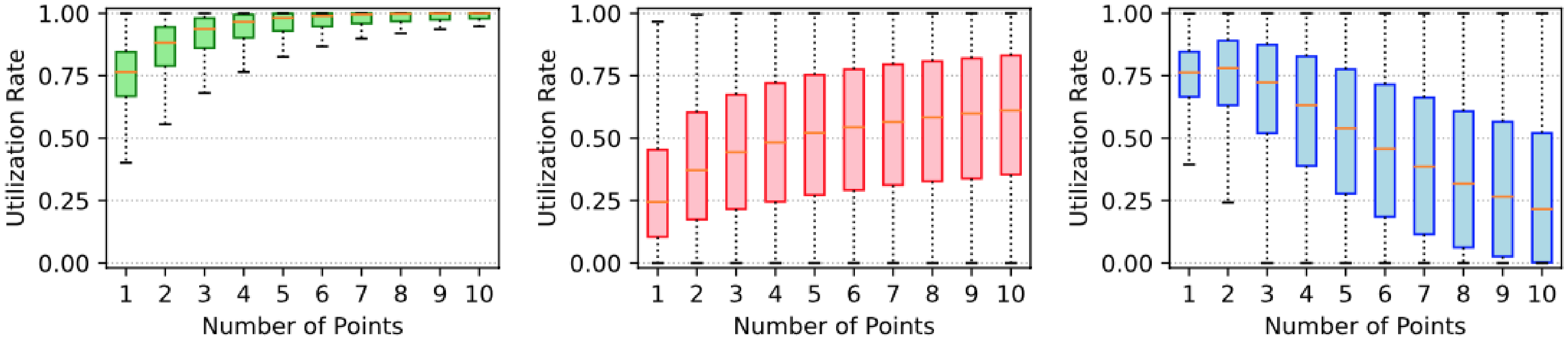
What's the Attack success rate in one-time obfuscation and permanent obfuscation?



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.

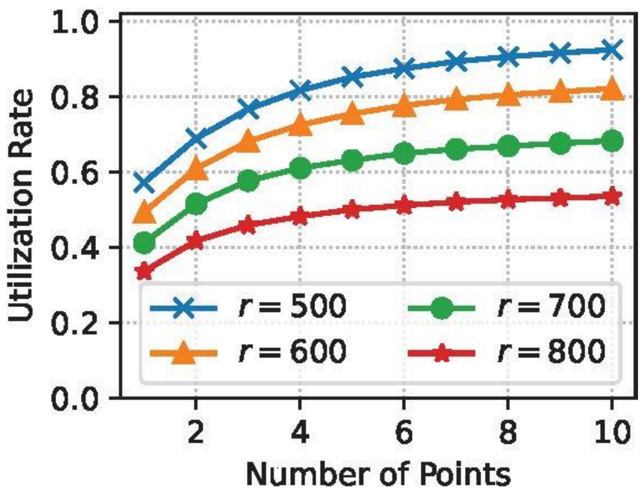
What's the performance of the n-fold Gaussian mechanism?



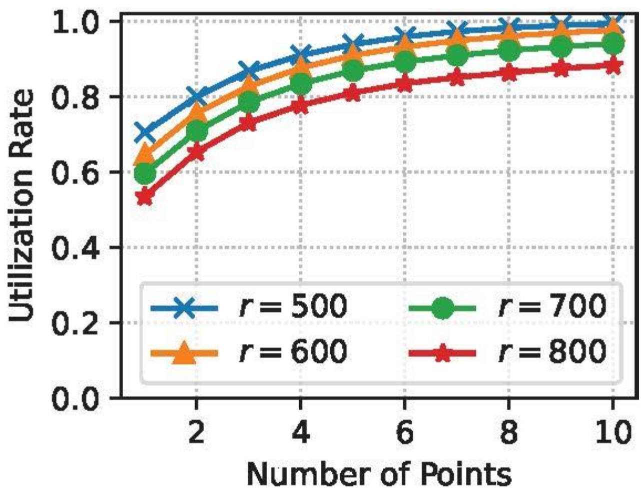
(a) n-fold Gaussian mechanism. (b) Post-processing mechanism. (c) Plain DP composition.

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?



$\varepsilon = 1$



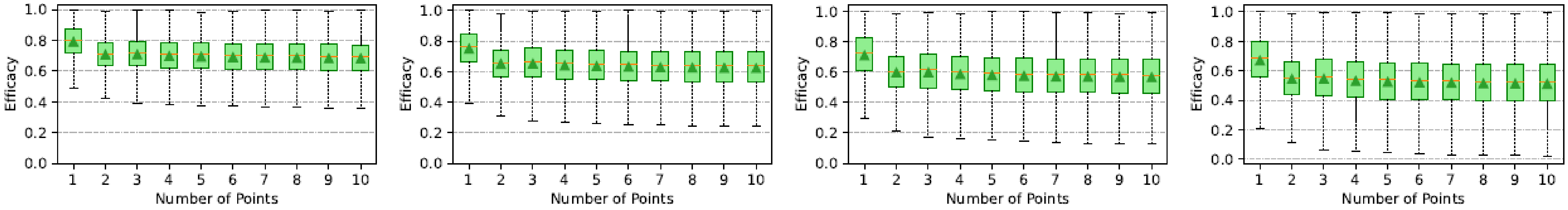
$\varepsilon = 1.5$

Observation 3

The utilization rate increase with n

Parameters: $\varepsilon = 1 \ or \ 1.5, \delta = 0.01$

What's the efficacy of Edge-PrivLocAd?



(a) r = 500.   (b) r = 600.   (c) r = 700.   (d) r = 800.

Observation 4
The efficacy do not significantly decrease with n
Parameters: $\varepsilon = 1, \delta = 0.01$

Scalability of Edge-PrivLocAd

- Emulation with Raspberry Pi 3

TABLE II: Obfuscation processing time.

Number of Users	2000	4000	8000	16000	32000
Processing Time (s)	340	627	1166	2090	4014

TABLE III: Output selection time.

Number of Users	2000	4000	8000	16000	32000
Processing Time (ms)	90	175	350	698	1377

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.