Mechanisms for Co-Location Privacy

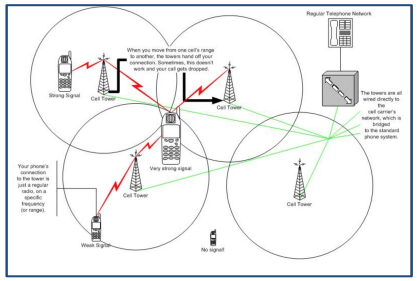
Ottilingam, Nithin Krishna

A Master's thesis

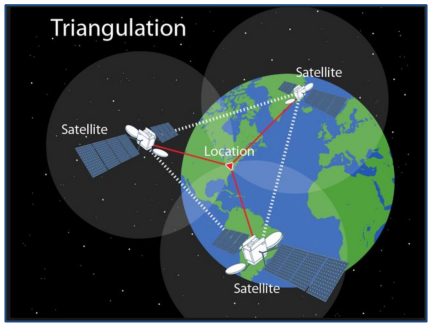
Submitted in part fulfilment of the requirements for the degree of,
Masters in Science in Computer Science, 2017.

Ubiquitous Location Data

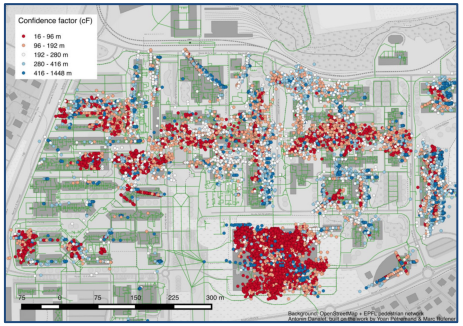
a) Systems record user's movements



Cellular location identification

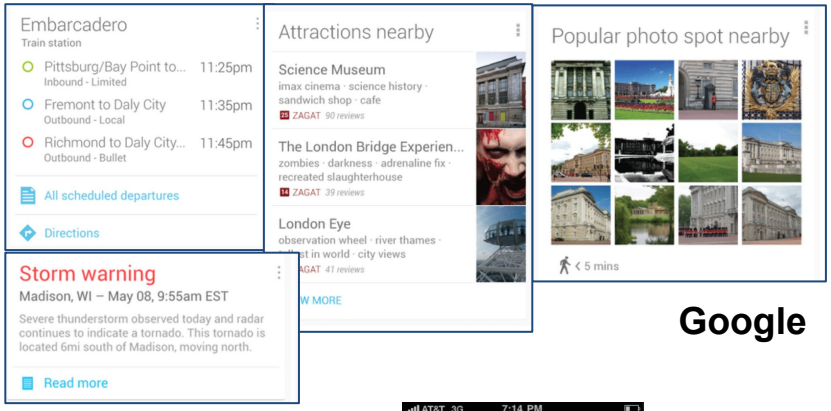


GPS systems



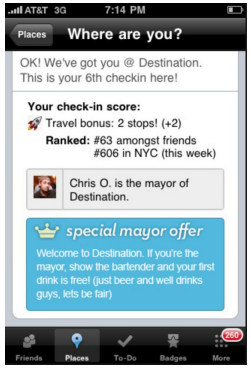
Wi-Fi Localization

b) Powerful **Location Based Services**(LBS)



Google

Facebook



4Square

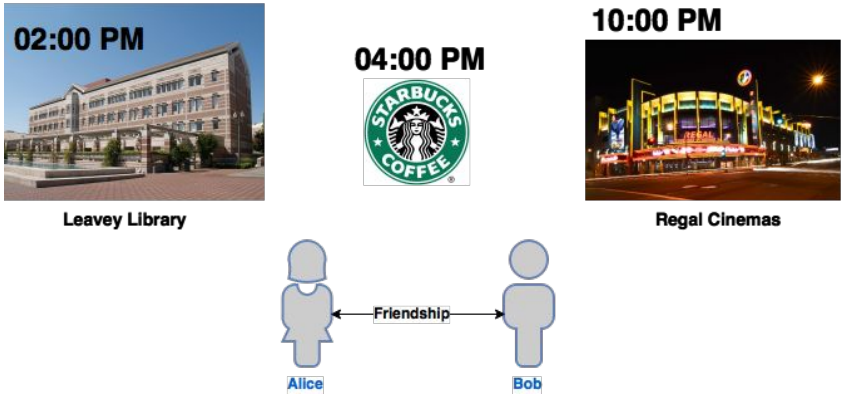
Motivation

Location data is necessary for service

Associated leakages with location data, not just location privacy

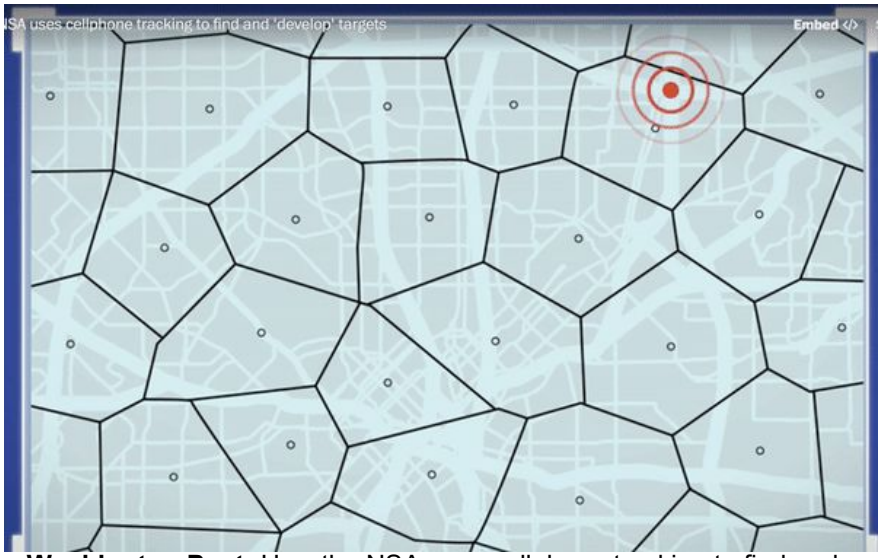
Co-Location Privacy Risks

Academia:
Co-location reveals social-ties between users



Studies predict future user location based on inferred social ties.

Realworld:
NSA's Co-Traveller program, Identifies unknown associates of a known target.



Washington Post: How the NSA uses cellphone tracking to find and 'develop' targets.

Co-Location Privacy Example

Every Fitbit User

Health & Location Data

fitbit

Health & Location Data

YOU

Everyone's Data

(Company X) Heart Attack Prediction

Prevent

Your social network

Emergency Response

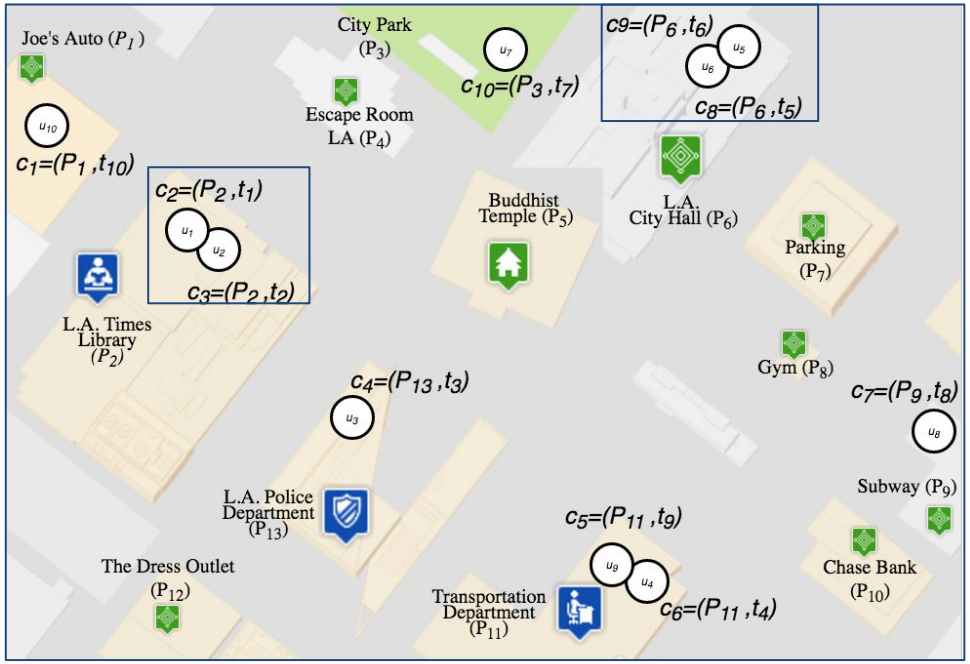
OVERVIEW

Location Data, Motivation, Co-Location Privacy Risks

- What is a Co-Location?
- Data setting and System Model
- Assumptions
- Co-Location Privacy methods
- Results

What is a Co-Location?

Co-Location: Two people at *roughly* the same geographic locale at roughly the same time.



We quantify 'roughly' based on parameters Δ_s and Δ_t .

- *Assume buildings are points*

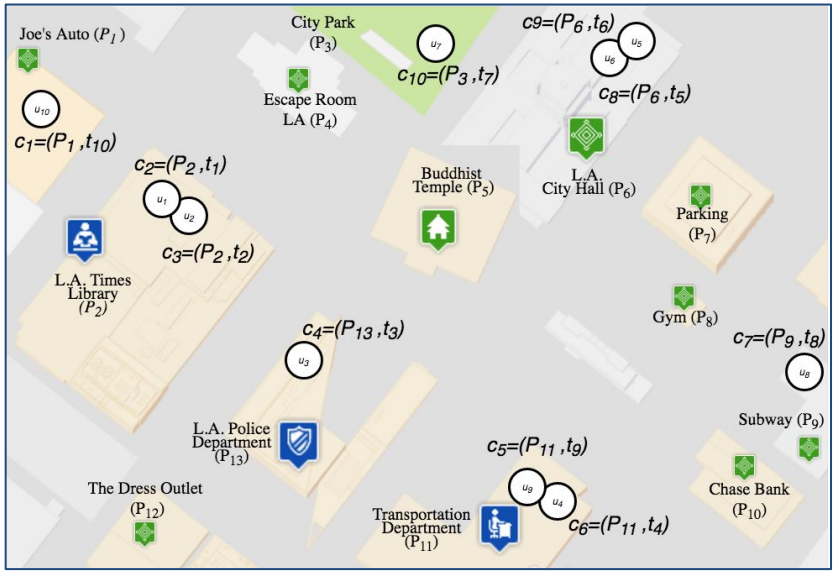
$\Delta_s = SameBuilding, \Delta_t = 1t$

Co-Locations: $(u_1, u_2), (u_5, u_6)$

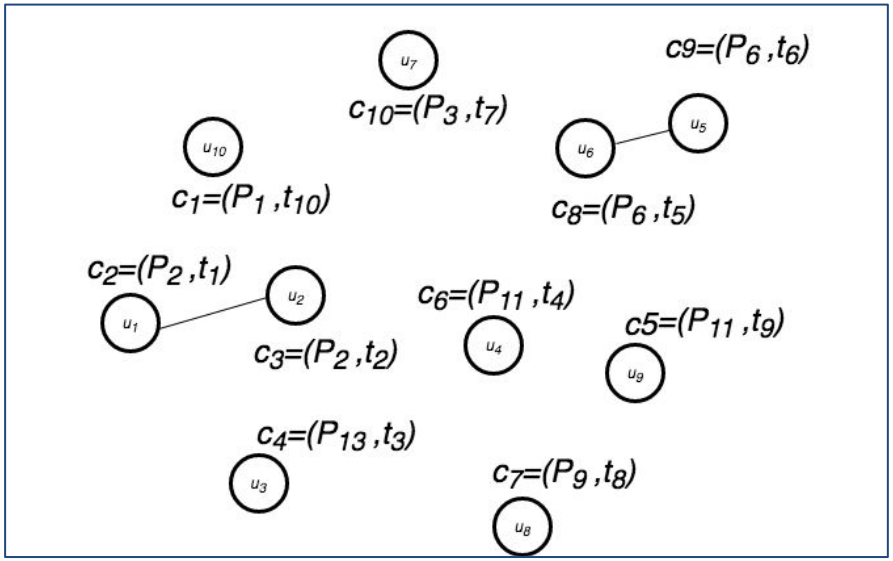
Note: Δ_s and Δ_t are application specific.

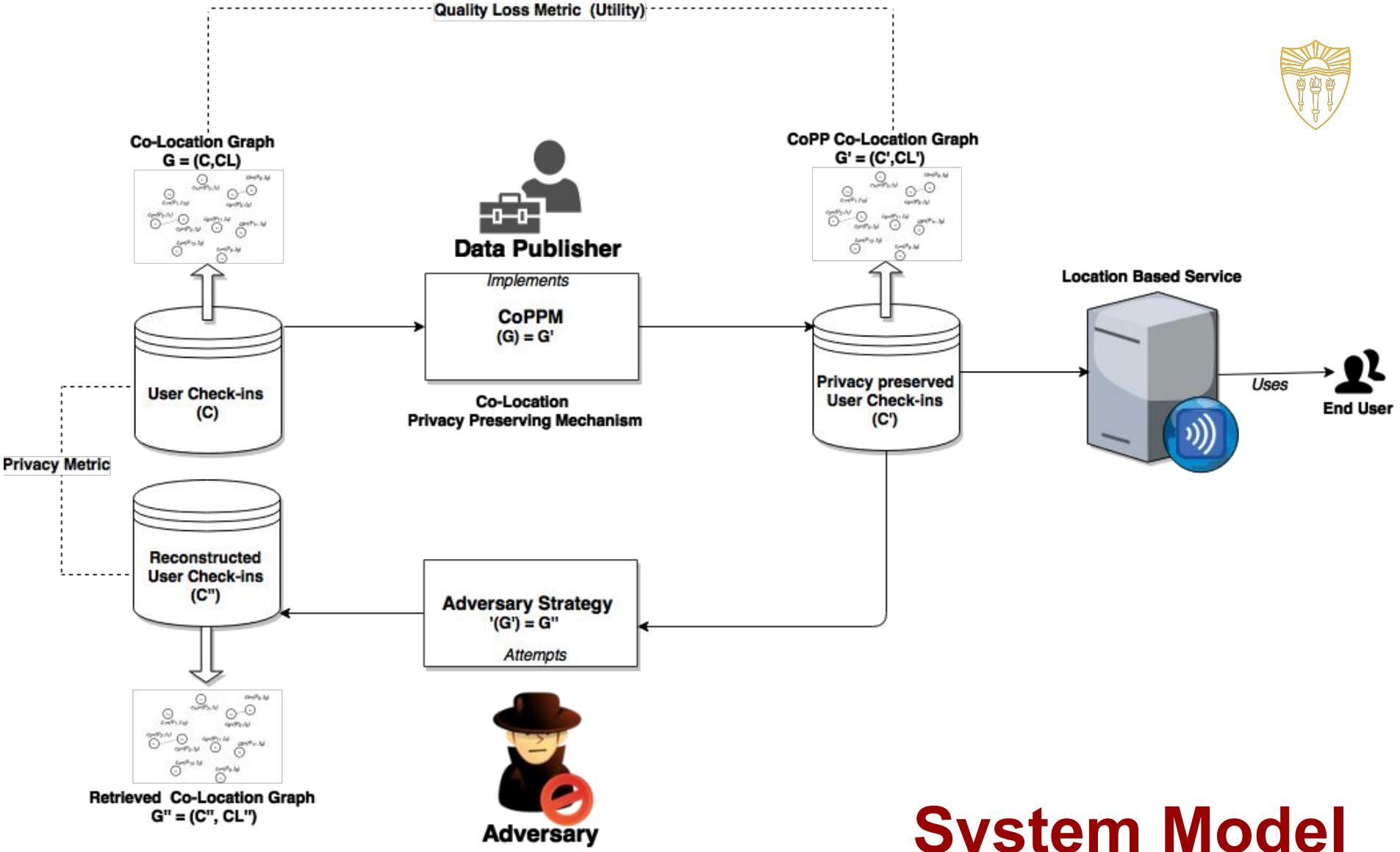
Data Setting

User Check-ins (C)



Co-Location Graph (G)





System Model

CoPPM: Co-Location Privacy Preserving Mechanism

Utility vs Privacy

Quality Loss Metric:
Linear combination of *spatial and temporal* distance between between, original and published check-ins.

$$QL = \text{Spatial Distortion} / MAX\_s + \text{Temporal Distortion} / MAX\_t$$

Total deterioration
Captures the ***Utility*** *of Location Based Services(LBS).*

Privacy Metric:
The accuracy of adversary's co-location reconstruction.

$$\text{Inference Accuracy } (\mathbf{IA}) = (C' \cap C) / C'$$

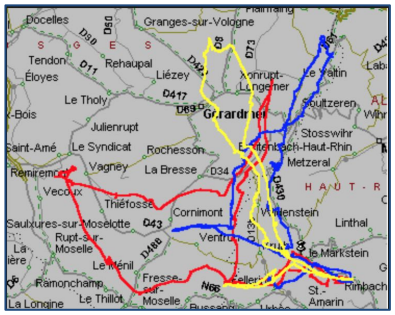
$$\text{Inference Recall } (\mathbf{IR}) = (C' \cap C) / C$$

Note: MAX_s and MAX_t are normalizing parameters, set by the data publisher.

Assumptions

1. Sporadic location exposures → No temporal correlation

Sporadic exposures



Continuous exposures

2. Set of Historical co-location between user pairs, eg. Couple

3. Every co-location is equally important

4. Preliminary Studies indicate, Several such correlations exist. Assume adversary remains unaware.

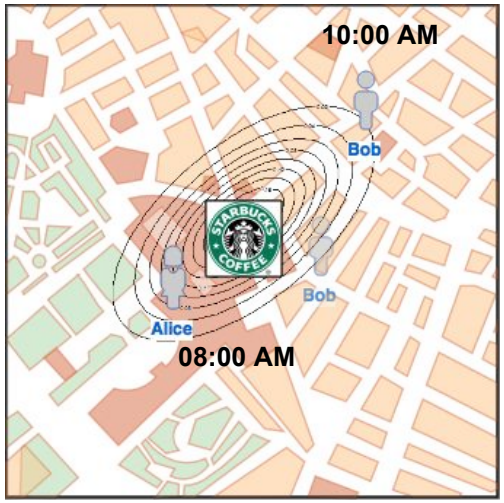
Co-Location Privacy Preserving Mechanisms

Method 1: Naïve Gaussian Perturbation

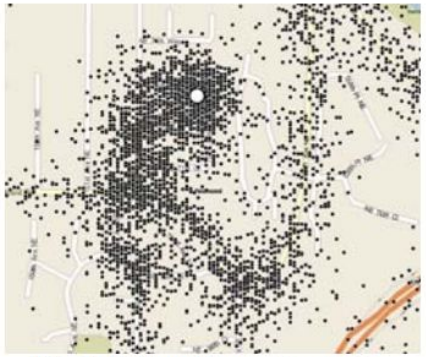
Simplest method in Location Privacy

Most popular methods in statistical data privacy

Method: For every co-location, it is enough to perturb one check-in.
Translate both coordinates with 2d-gaussian noise.
Translate timestamp with 1d-gaussian noise



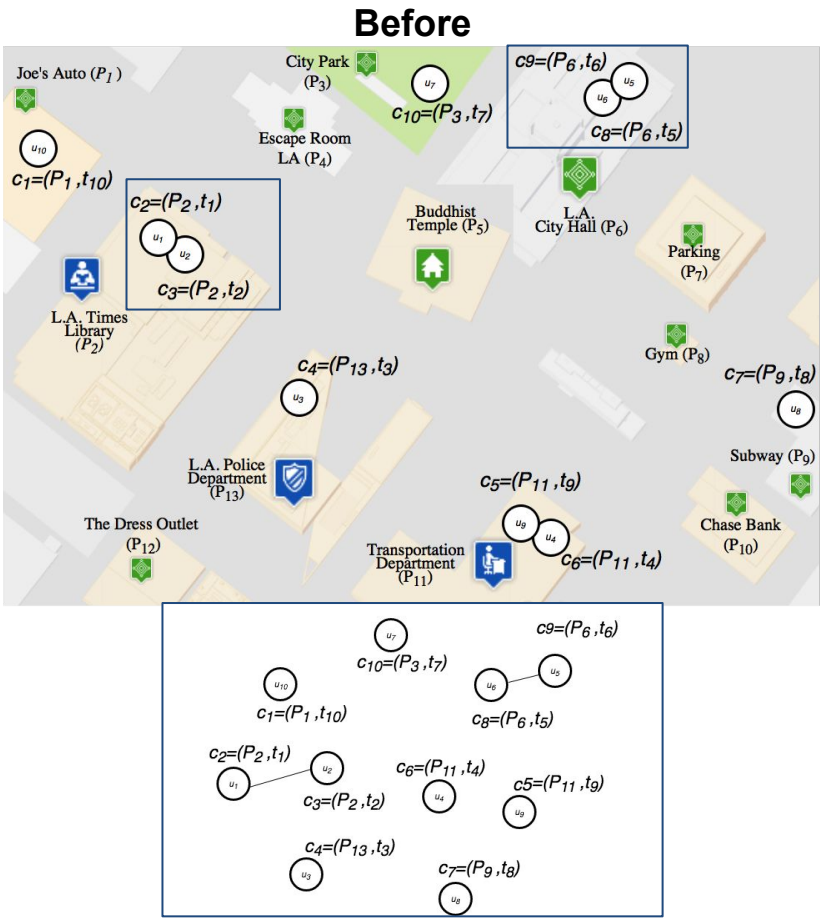
(a) Original GPS data

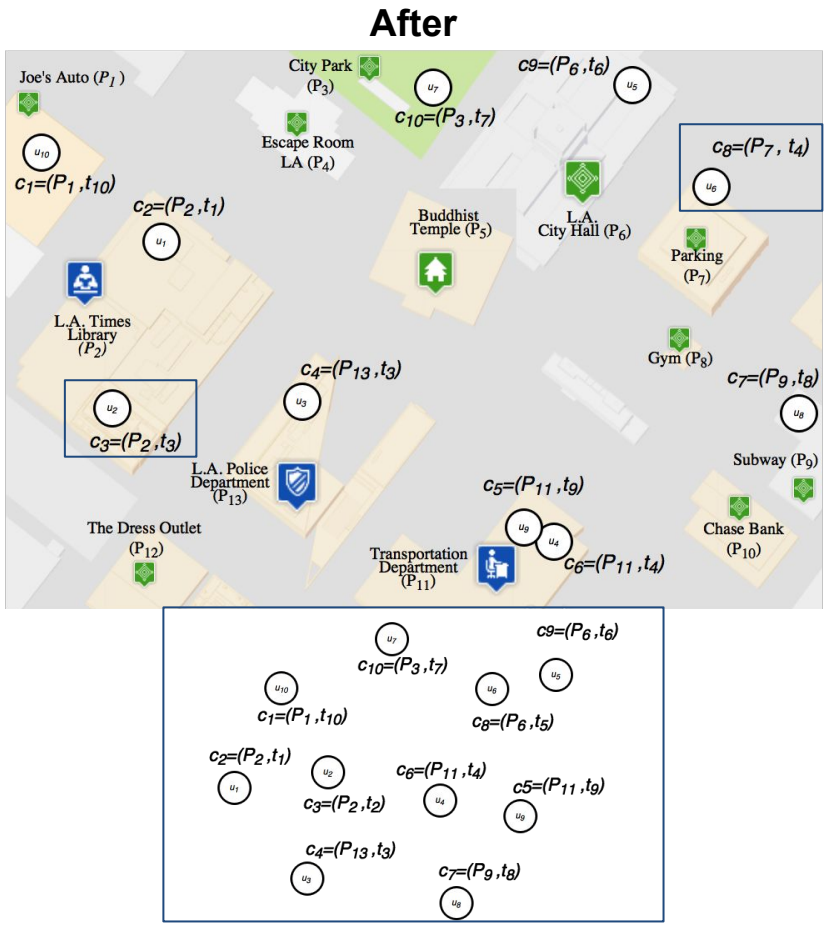


(b) Additive Gaussian noise

Inference attacks on Location Traces, John Krumm

Naïve Gaussian Perturbation

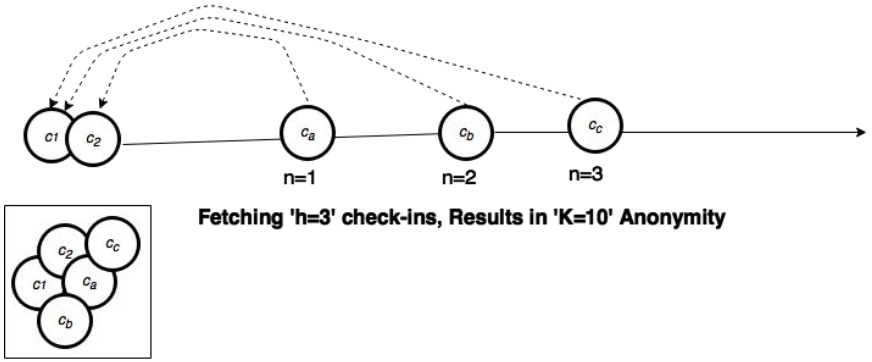




Method 2: Co-Location K-Anonymity

Definition 1. *Given a co-location network $G = (C, CL)$, a co-location $cl \in CL$ is said to be k-anonymous if it is spatio-temporally indistinguishable to $k-1$ other co-locations.*

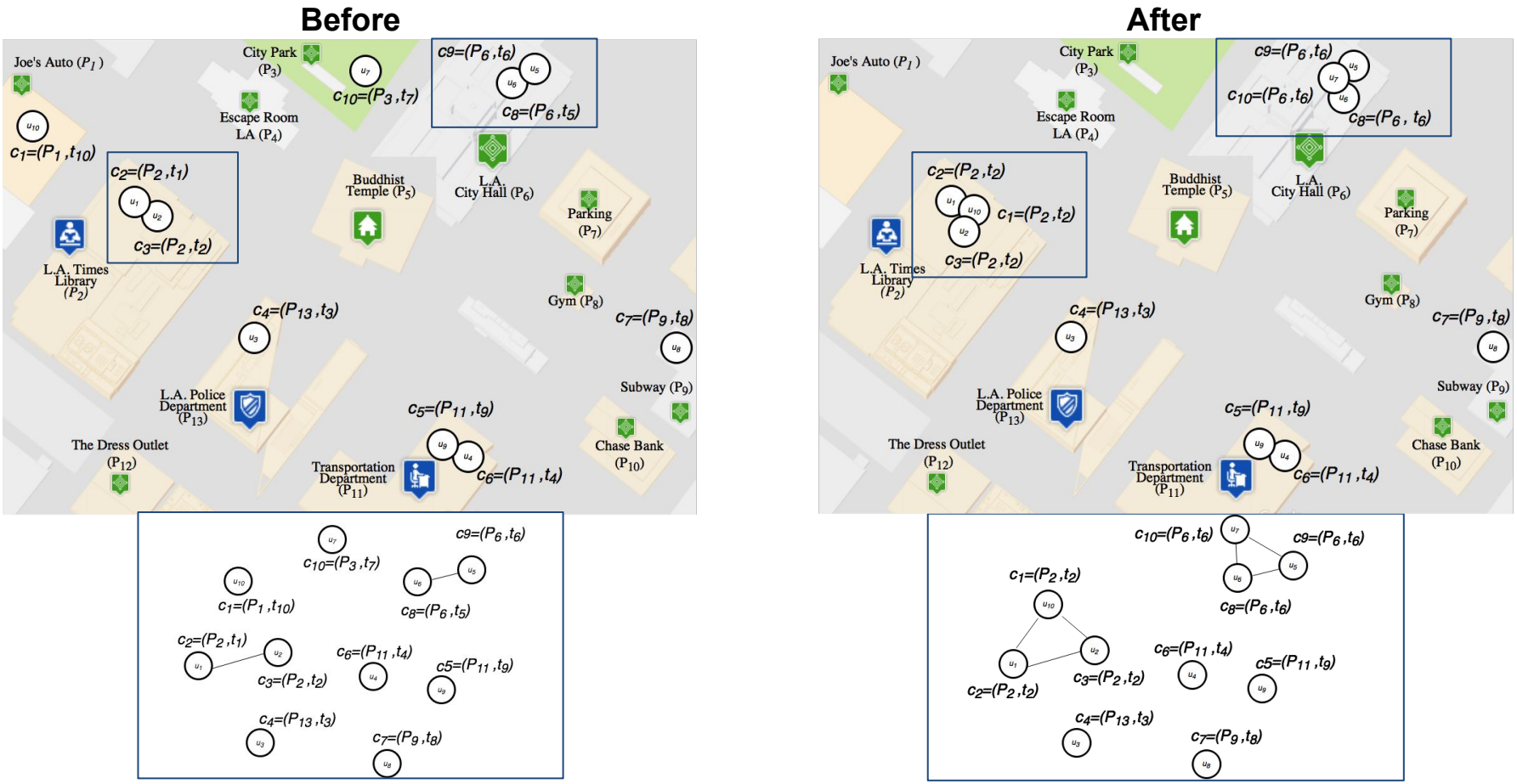
Method: For every co-location pair, Make each co-location k-anonymous by moving "h" closest check-ins to form a group.



Note: We define ***closest*** based on a linear combination of Spatial and Temporal distances.

$$ST(c,c') = \Sigma\ \alpha/MAX_s .|| c, c' ||_s + (1-\alpha)/MAX_t .|| c, c' ||_t$$

Co-Location K-Anonymity

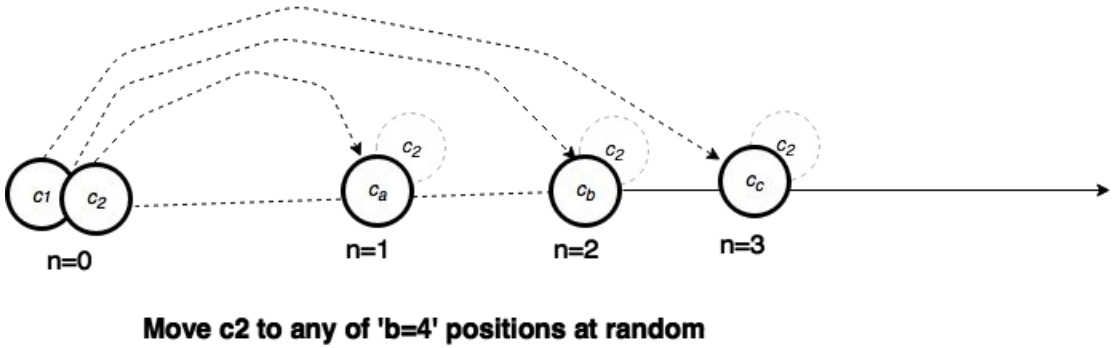


Method 3: Adaptive Perturbation

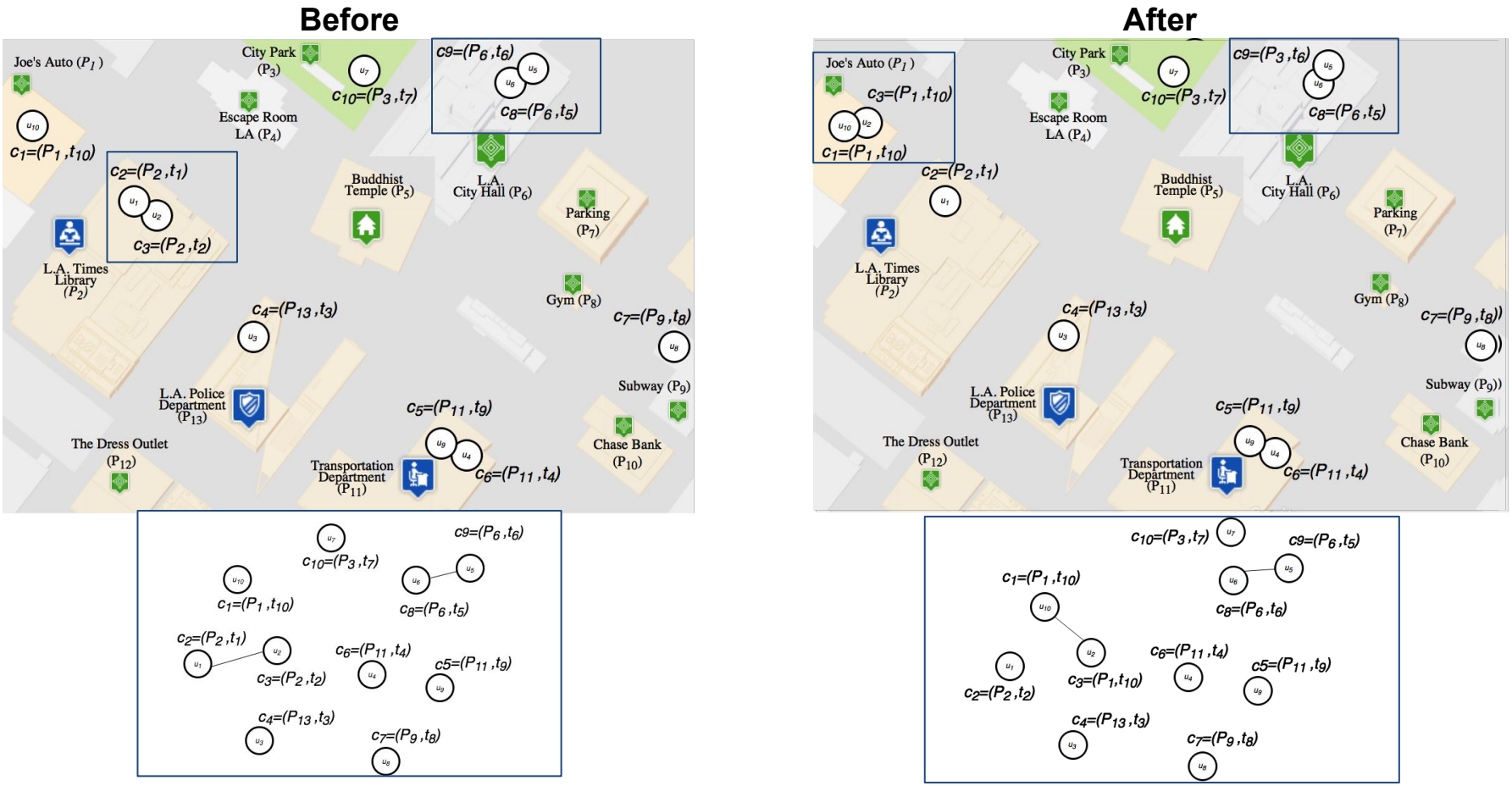
K-Anonymity may lead to excessive Quality Loss
(eg) To make a co-location 4-anonymous, it results in the perturbation of upto 4 check-ins.

Can we do better?

Method: For every co-location pair, pick one check-in at random;
Move it to one of it's $[0, b\text{-}1]$th nearest neighbours,
with a probability $1/b$.



Adaptive Perturbation



RESULTS

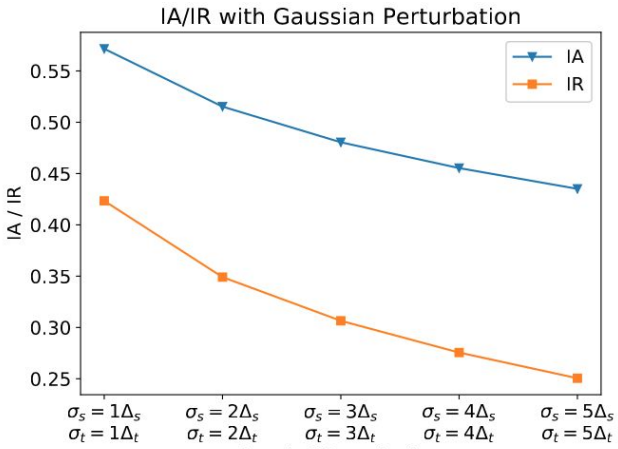
Dataset & Parameter Setting

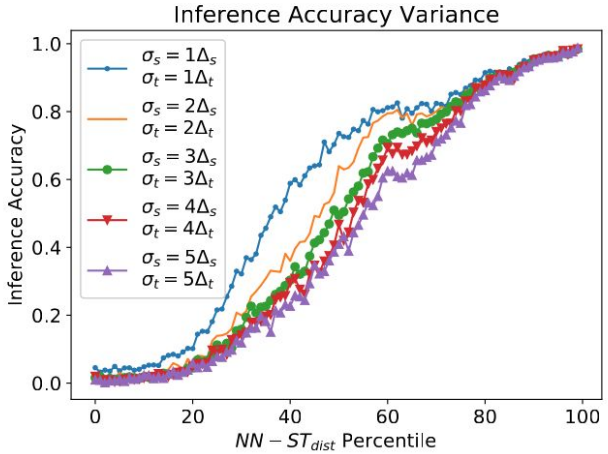
Dataset	Gowalla-Small	Gowalla-Big	Shanghai
Check-ins	3.6m	18m	7.5m
Users	55k	120k	270k
Locations	600k, in the US	1.2M, in the US	45k
Time-Period	18 months	4 years	1 year
Check-ins Per Hour (City Scale)	38	95	865
Check-ins Per Sq.Mile (City Scale)	1200	5800	9000

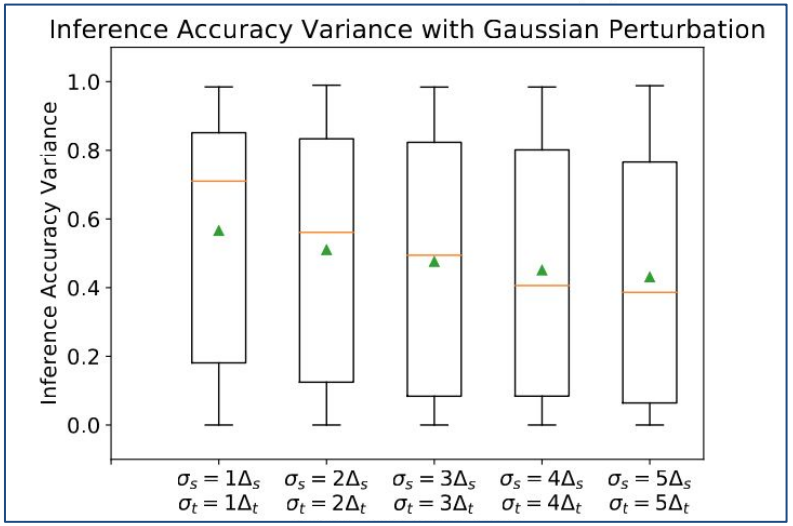
Parameters:

1. Δ_s = 25 meters and Δ_t = 20 minutes
2. α = 0.5
3. MAX_s = 5km and MAX_t = 48hours: discard checkins in extremely sparse regions.

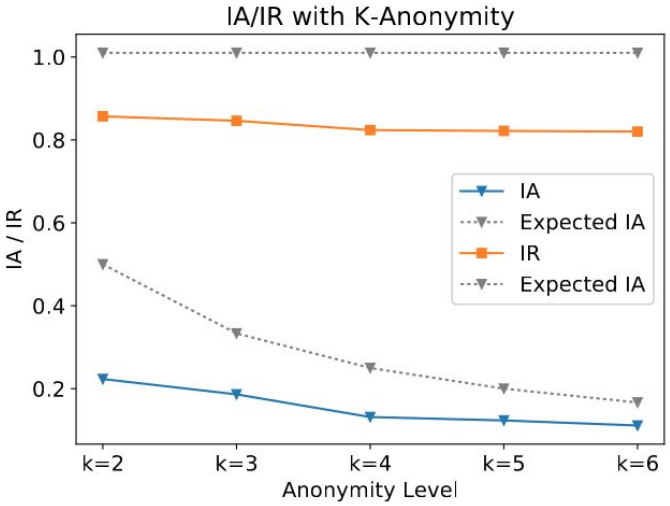
Baseline

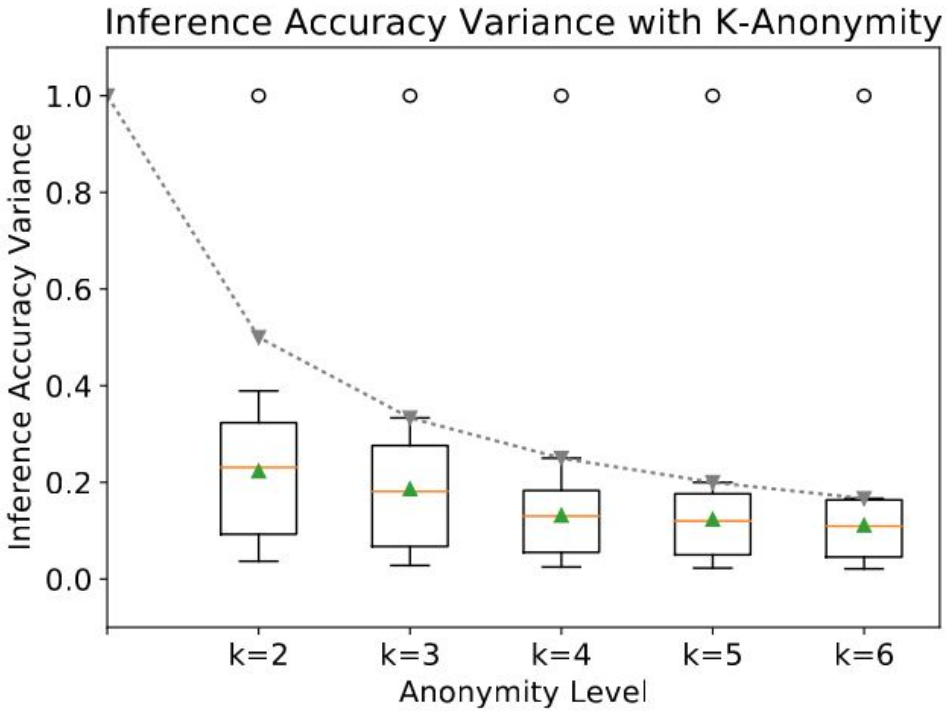

IA/IR with Gaussian Perturbation
IA
IR
IA / IR
0.55
0.50
0.45
0.40
0.35
0.30
0.25
$\sigma_s = 1\Delta_s$ $\sigma_t = 1\Delta_t$
$\sigma_s = 2\Delta_s$ $\sigma_t = 2\Delta_t$
$\sigma_s = 3\Delta_s$ $\sigma_t = 3\Delta_t$
$\sigma_s = 4\Delta_s$ $\sigma_t = 4\Delta_t$
$\sigma_s = 5\Delta_s$ $\sigma_t = 5\Delta_t$


Inference Accuracy Variance
Inference Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
$NN - ST_{dist}$ Percentile
$\sigma_s = 1\Delta_s$ $\sigma_t = 1\Delta_t$
$\sigma_s = 2\Delta_s$ $\sigma_t = 2\Delta_t$
$\sigma_s = 3\Delta_s$ $\sigma_t = 3\Delta_t$
$\sigma_s = 4\Delta_s$ $\sigma_t = 4\Delta_t$
$\sigma_s = 5\Delta_s$ $\sigma_t = 5\Delta_t$

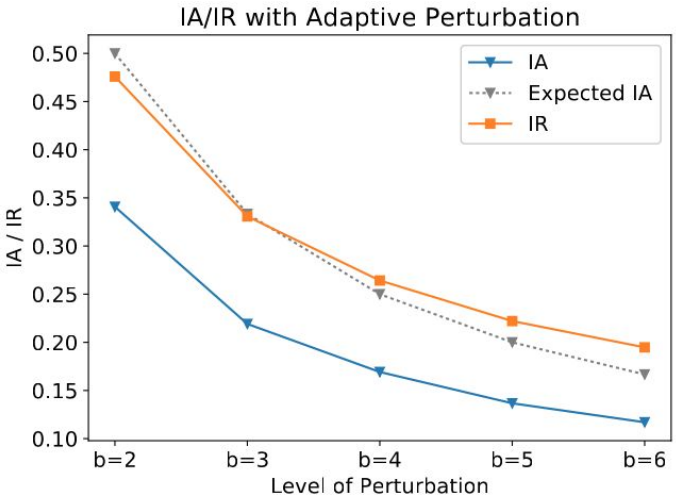

Inference Accuracy Variance with Gaussian Perturbation
Inference Accuracy Variance
1.0
0.8
0.6
0.4
0.2
0.0
$\sigma_s = 1\Delta_s$ $\sigma_t = 1\Delta_t$
$\sigma_s = 2\Delta_s$ $\sigma_t = 2\Delta_t$
$\sigma_s = 3\Delta_s$ $\sigma_t = 3\Delta_t$
$\sigma_s = 4\Delta_s$ $\sigma_t = 4\Delta_t$
$\sigma_s = 5\Delta_s$ $\sigma_t = 5\Delta_t$

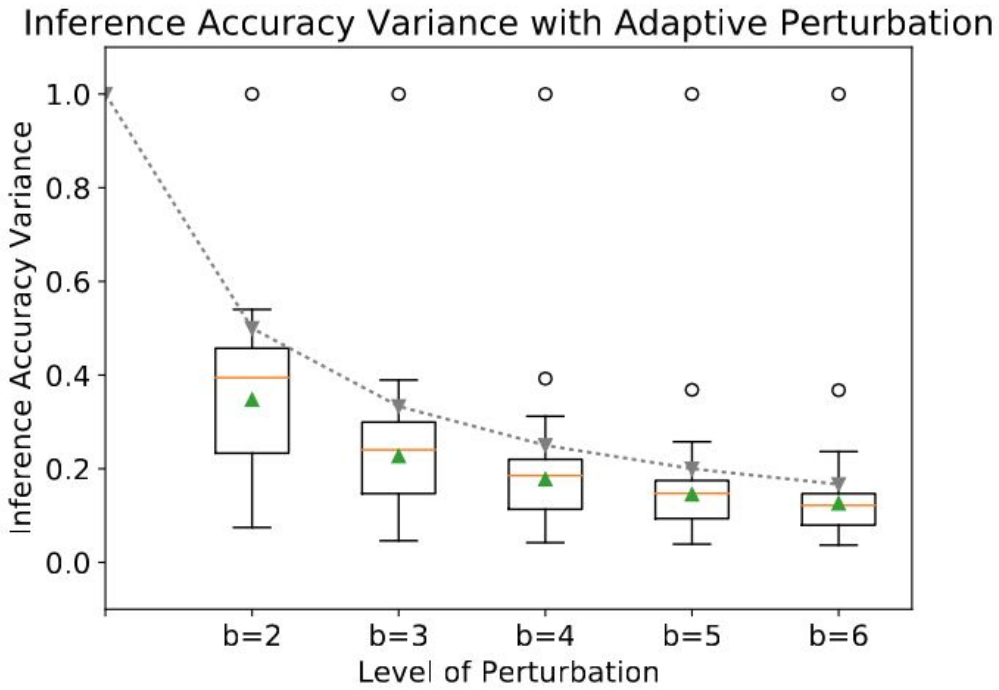
K-Anonymity


IA/IR with K-Anonymity
IA / IR
1.0
0.8
0.6
0.4
0.2
IA
Expected IA
IR
Expected IA
k=2
k=3
k=4
k=5
k=6
Anonymity Level

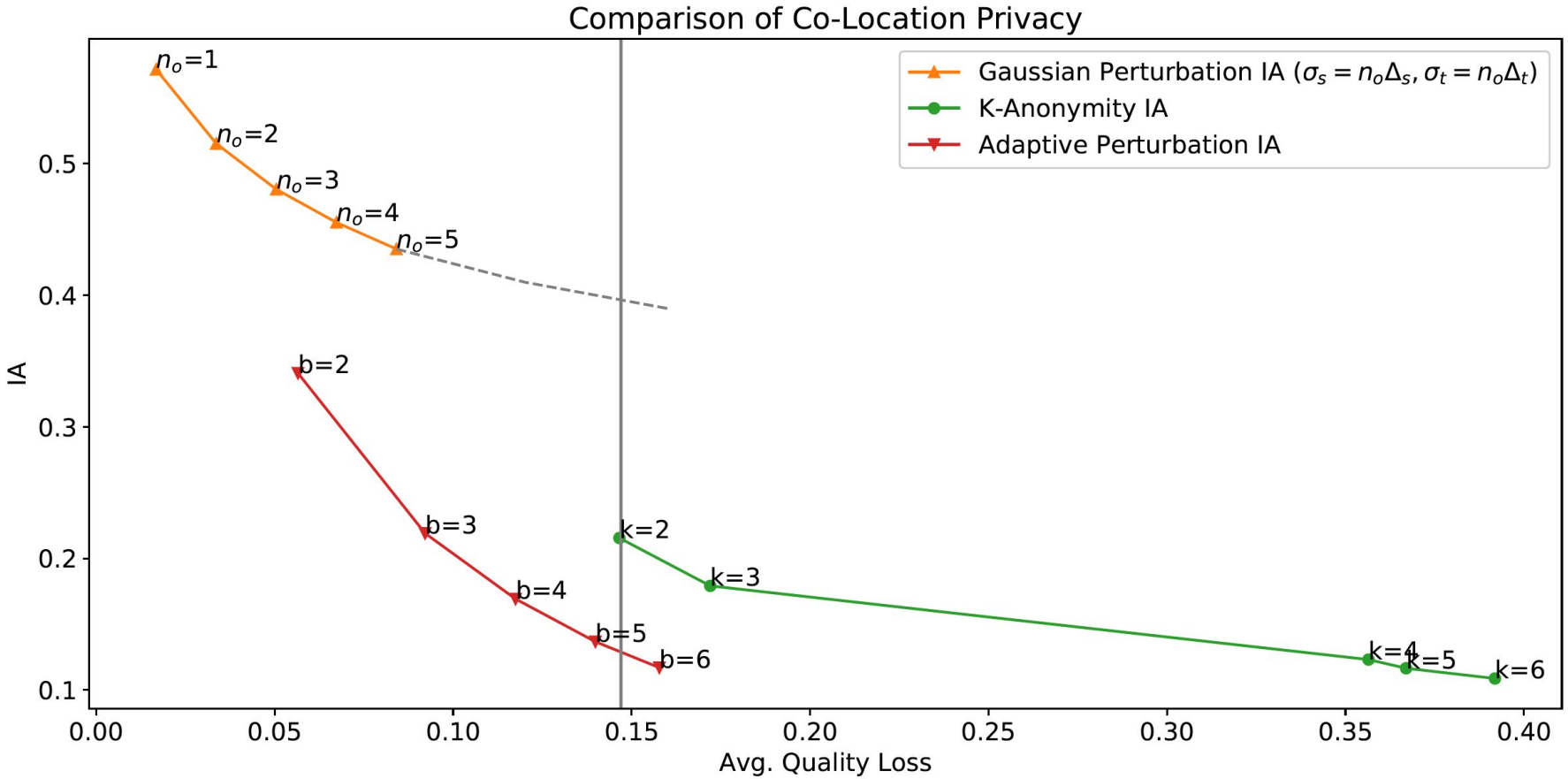

Inference Accuracy Variance with K-Anonymity
Inference Accuracy Variance
1.0
0.8
0.6
0.4
0.2
0.0
k=2
k=3
k=4
k=5
k=6
Anonymity Level

Adaptive Perturbation





Between Methods



LBS Utility - ST Range

CONCLUSION

1) Introduced the problem of co-location privacy

2) Proposed some interesting methods, evaluated them

3) Note that proposed methods not optimal.
   Susceptible to a knowledgeable adversary.

4) Future look to expand this work and address:
   a) Live location updates
   b) Location traces

Thank You, Questions?