IEEE International Conference on Mobile Data Management
Workshops/PhD Forum, 06/10/2019

Hide in Plain Sight: Enabling Mobile Applications and Data Analytics with Local Differential Privacy

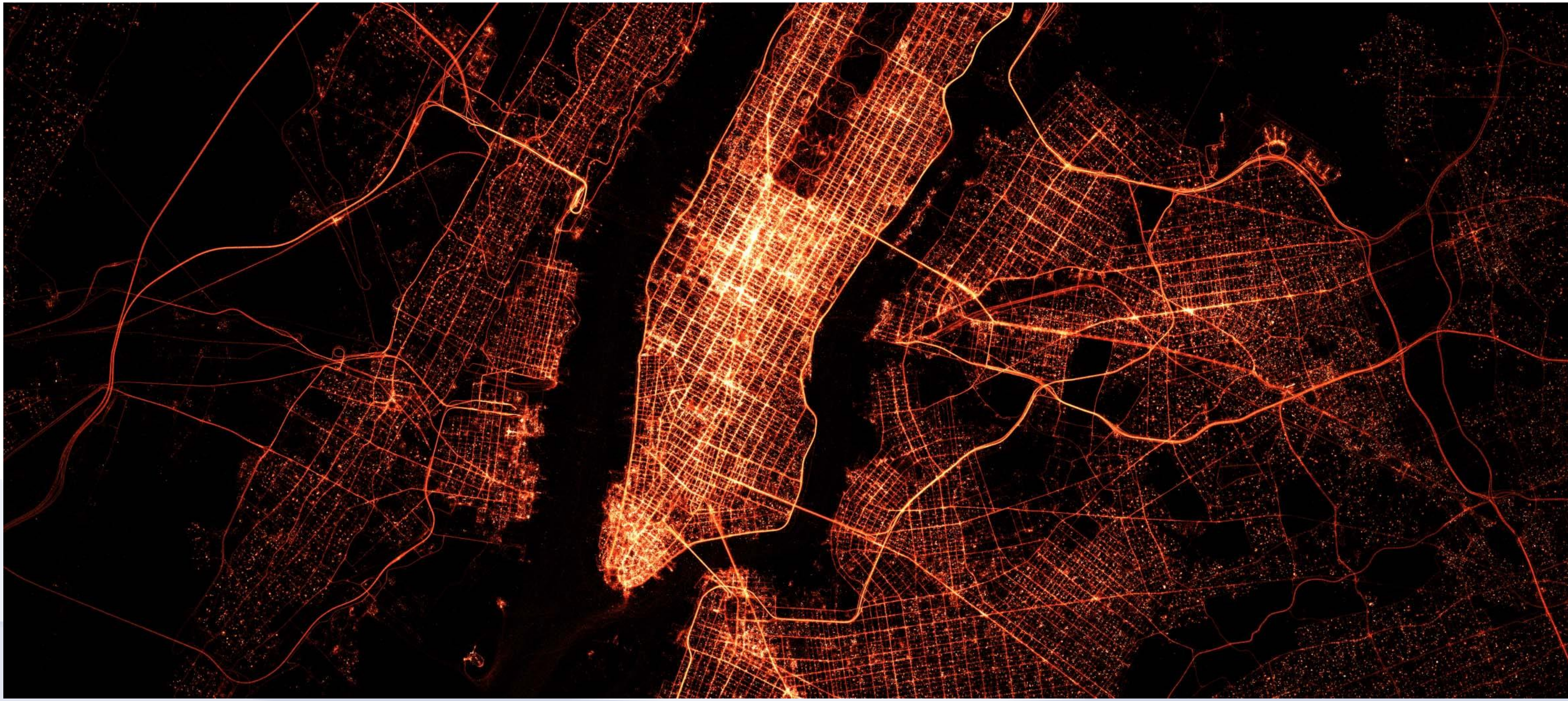
Li Xiong
Department of Computer Science
Department of Biomedical Informatics
Emory University

* Work supported by National Science Foundation and Google Research Award

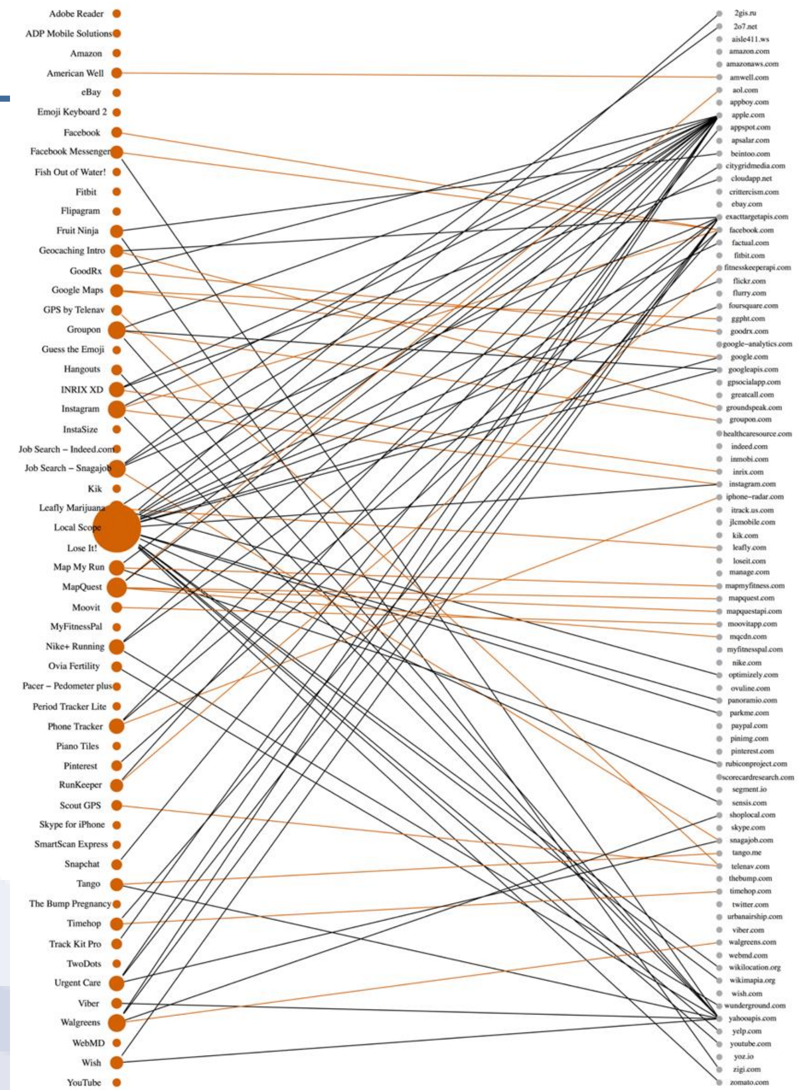
EMORY
UNIVERSITY

Location data collected from individual devices
(Source: New York Times 12/2018)

EMORY UNIVERSITY



Over 235 million locations captured from more than 1.2 million unique devices during a three-day period in 2017 (Source: New York Times 12/2018)

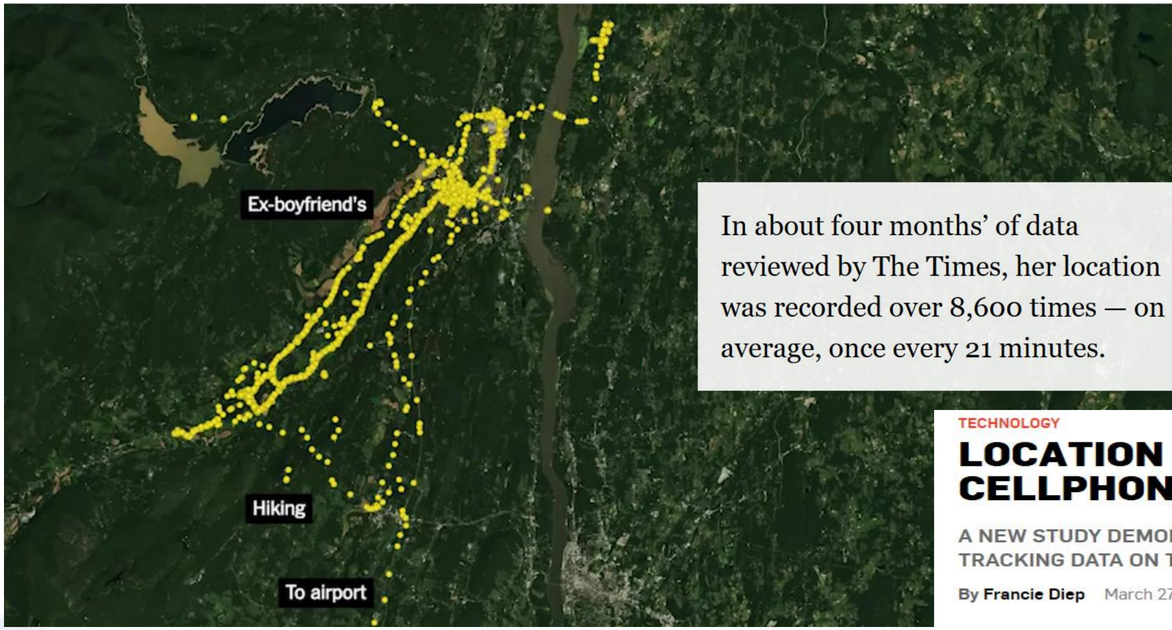


Location data sharing by iOS apps (left) to domains (right)

33%/47% of **Android/ iOS** apps shared GPS coordinates with third parties

Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps, 2015-10-30 https://techscience.org/a/2015103001/

EMORY UNIVERSITY

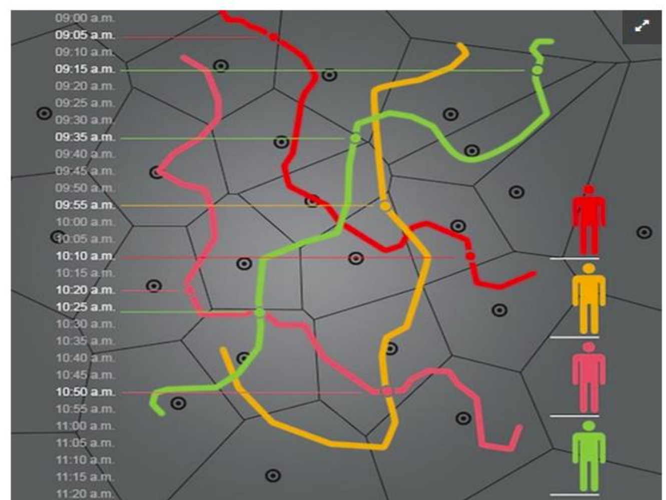


TECHNOLOGY

LOCATION DATA CAN UNIQUELY IDENTIFY CELLPHONE USERS

A NEW STUDY DEMONSTRATES HOW EASY IT IS TO IDENTIFY PEOPLE FROM THE LOCATI
TRACKING DATA ON THEIR CELLPHONES.

By Francie Diep March 27, 2013



Related Content

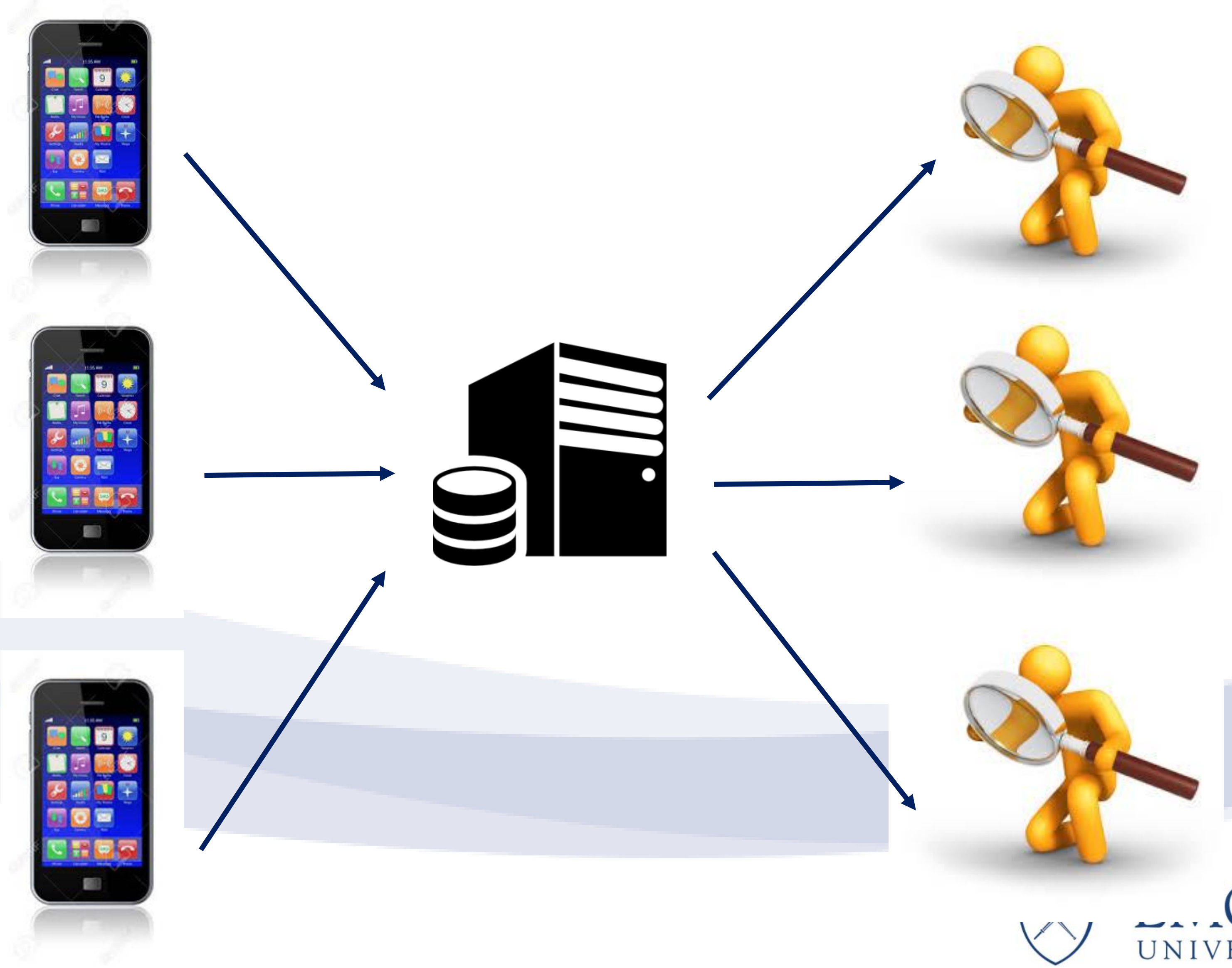
Text Message
Ukrainian Pro
Cellphones Ar
Tracked

New Tracking
Pin Any Intern
Location to Wi
Hundred Mete

How Stores Tr
Shoppers

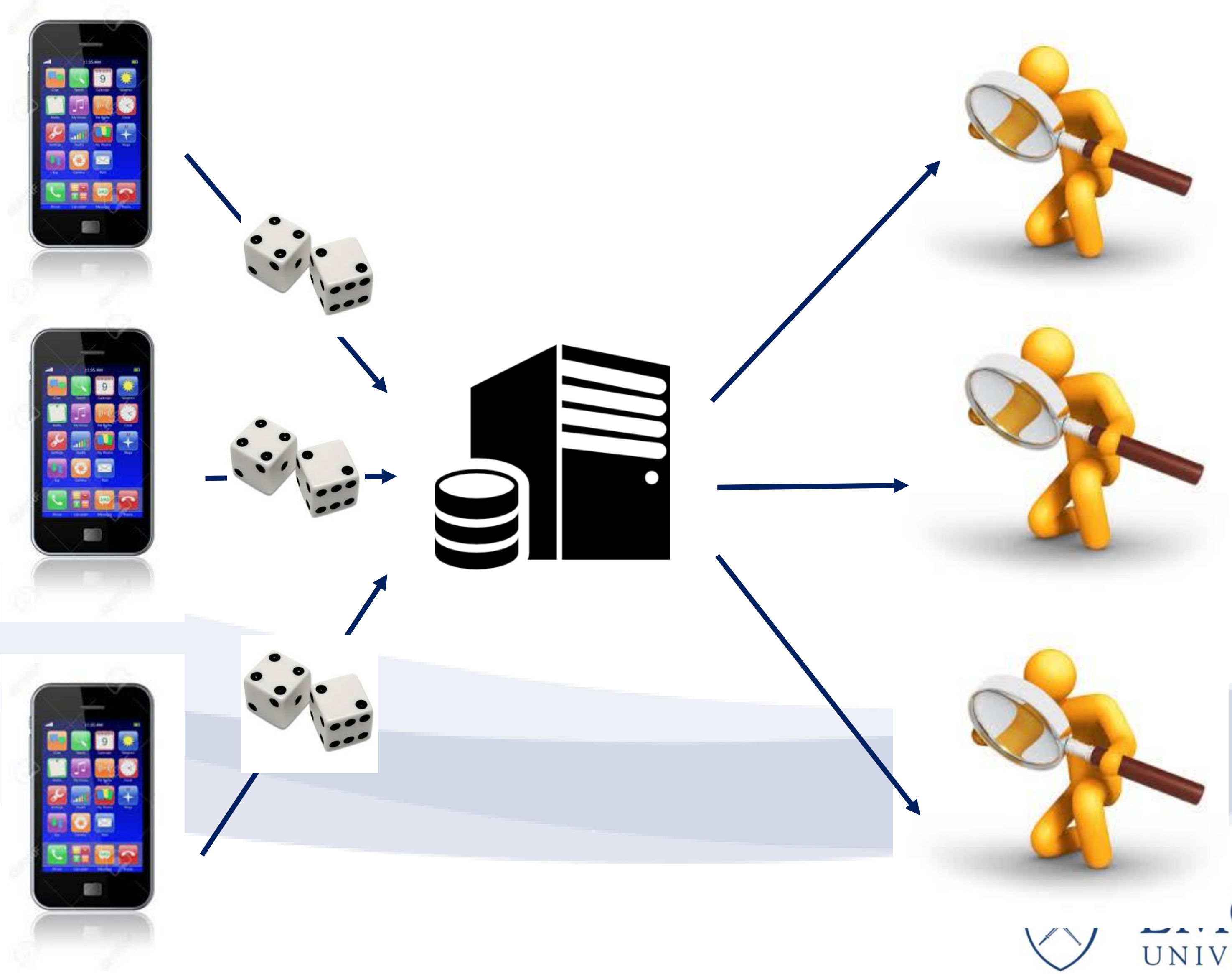
Facebook Will
Physical Store

The Mobile Data Economy



Enabling Data Analytics with Centralized Differential Privacy

Enable Mobile Apps and Analytics with Local Differential Privacy



Enabling Mobile Apps and Analytics with Local Differential Privacy

- Background
  - Local differential privacy
  - Geo-indistinguishability (local d-privacy)
- Extended privacy notions
  - Protecting dynamic locations (CCS15, VLDB17 demo)
  - Protecting spatiotemporal events (ICDE19)
- New mobile applications
  - Spatial crowdsourcing with geo-indistinguishability (ICDE18)
- New mechanisms
  - Supporting both analytics and mobile applications (CNS19)

EMORY UNIVERSITY

Local Differential Privacy

- Privacy definition
  - Any two locations produce "similar' distributions (bounded by ϵ)

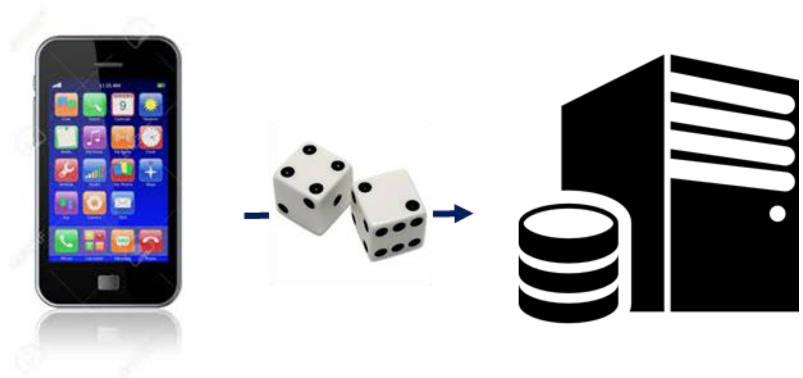
$$\frac{Pr(\mathcal{A}(\mathbf{x}_1) = \mathbf{z}_t)}{Pr(\mathcal{A}(\mathbf{x}_2) = \mathbf{z}_t)} \leq e^{\epsilon}$$

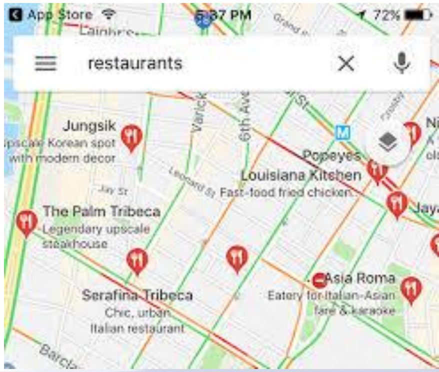
- Mechanism
  - Randomized response (with encoding)
- Applications
  - Simple analytics (e.g. frequency estimation)
  - Google, Apple, Microsoft
- Limitations
  - Output not useful for mobile apps

Geo-Indistinguishability (Local d-privacy)

- Privacy Definition
  - Any two locations at distance at most r produce “similar” distributions proportional to the distance (bounded by $\epsilon\, r$)

$$\frac{Pr(\mathcal{A}(\mathbf{x}_1) = \mathbf{z}_t)}{Pr(\mathcal{A}(\mathbf{x}_2) = \mathbf{z}_t)} \leq e^{\epsilon\, d(x1,x2)}$$



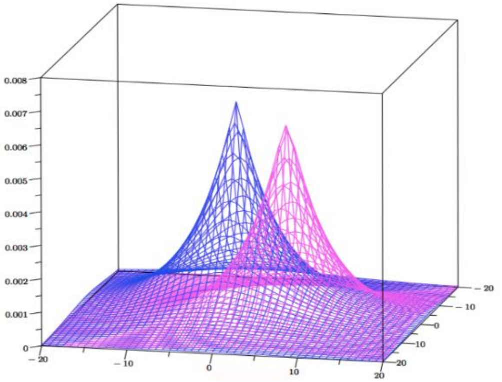


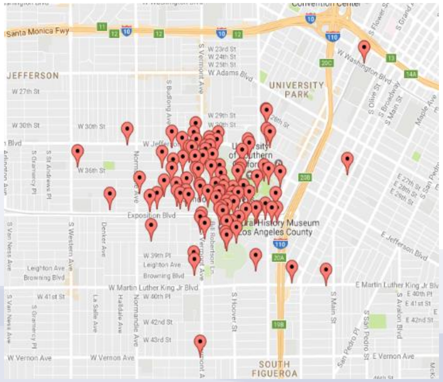
- Mechanism:
  - Planar Laplace mechanism
- Applications
  - Mobile apps/location sharing
- Limitations:
  - Temporal correlations of dynamic locations not considered
  - Not optimal for analytics

EMORY UNIVERSITY

Geo-Indistinguishability: Planar Laplace Mechanism

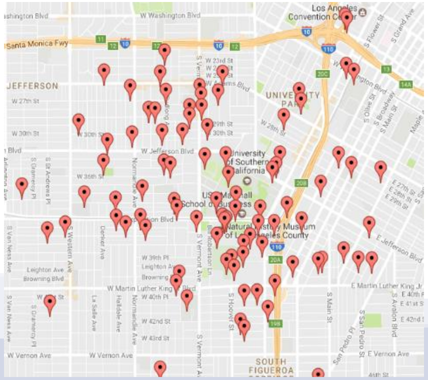
- Generating random point z (from actual point x ∈ X) according to planar Laplace distribution





$\epsilon = \log(6)$

r=1 km



Better privacy: $\epsilon = \log(2)$

r=1 km

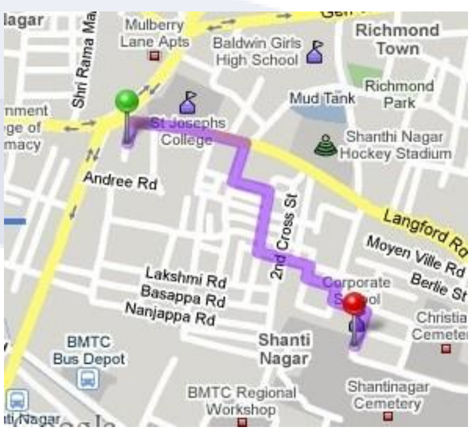
Enabling Mobile Apps and Analytics with Local Differential Privacy

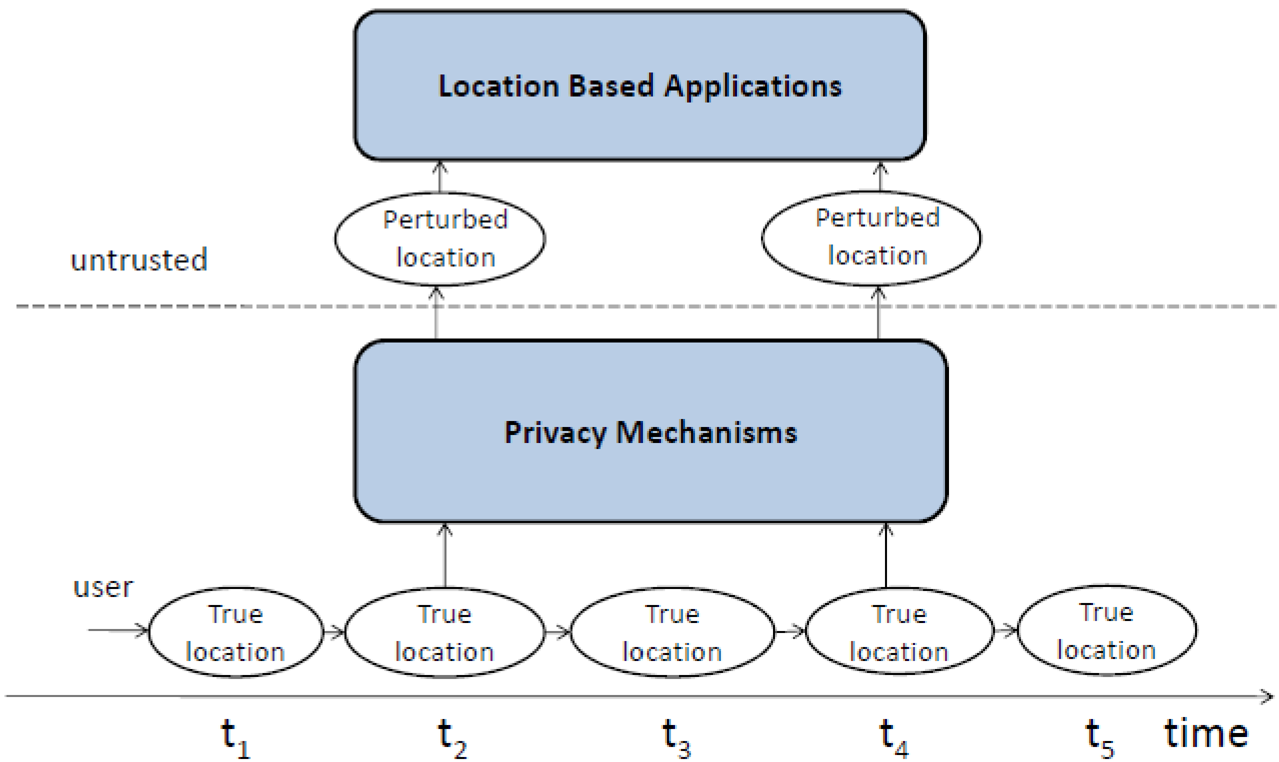
- Background
  - Local differential privacy
  - Geo-indistinguishability (local d-privacy)
- Extended privacy notions
  - Protecting dynamic locations (CCS15, VLDB17 demo)
  - Protecting spatiotemporal events (ICDE19)
- New mobile applications
  - Spatial crowdsourcing with geo-indistinguishability (ICDE18)
- New mechanisms
  - Supporting both analytics and mobile applications (CNS19)

EMORY UNIVERSITY

Location Privacy: Temporal Correlations

- Temporal correlations (adversary knowledge): moving patterns and previously released perturbed locations

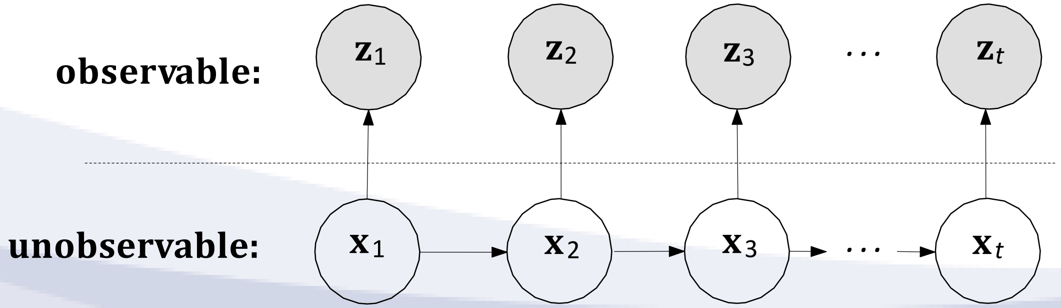




Differential Privacy with δ-location set

- δ-location set differential privacy
  - Any two locations in the probable location set produce "similar" distributions proportional to the distance (bounded by ϵ)
  - Probable location set determined by hidden Markov Model

$$\frac{Pr(\mathcal{A}(\mathbf{x}_1) = \mathbf{z}_t)}{Pr(\mathcal{A}(\mathbf{x}_2) = \mathbf{z}_t)} \leq e^{\epsilon}$$



Y. Xiao, L. Xiong. Protecting Locations with Differential Privacy under Temporal Correlations. CCS 2015
Y. Xiao, L. Xiong, S. Zhang, Y. Cao. LocLok: Location Cloaking with Differential Privacy via Hidden Markov Model. VLDB demo, 2017

EMORY UNIVERSITY

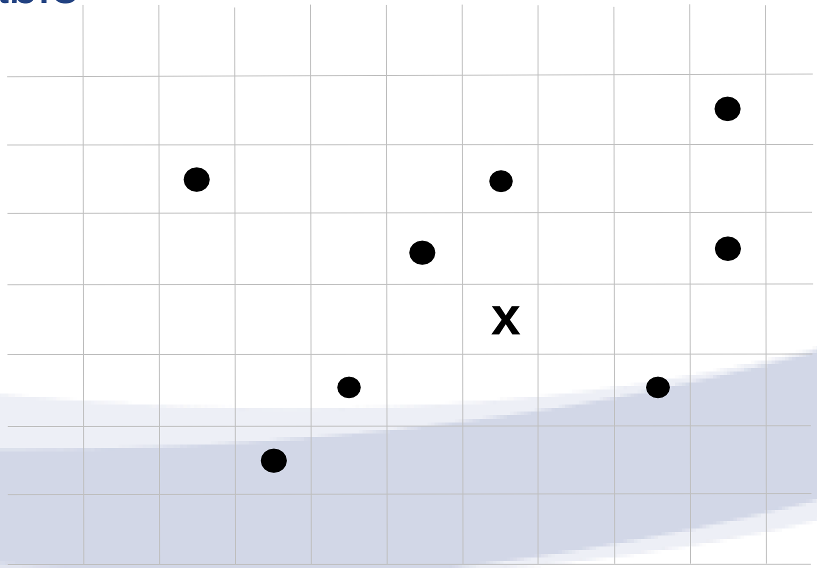
Optimal perturbation mechanism

- Minimize **expected distance** between perturbed location z and true location x

$$\textsc{Error} = \sqrt{\mathbb{E}||\mathbf{z} - \mathbf{x}^*||_2^2}$$

- While satisfying constraint of **differential privacy** – any pair of locations x1 and x2 are indistinguishable

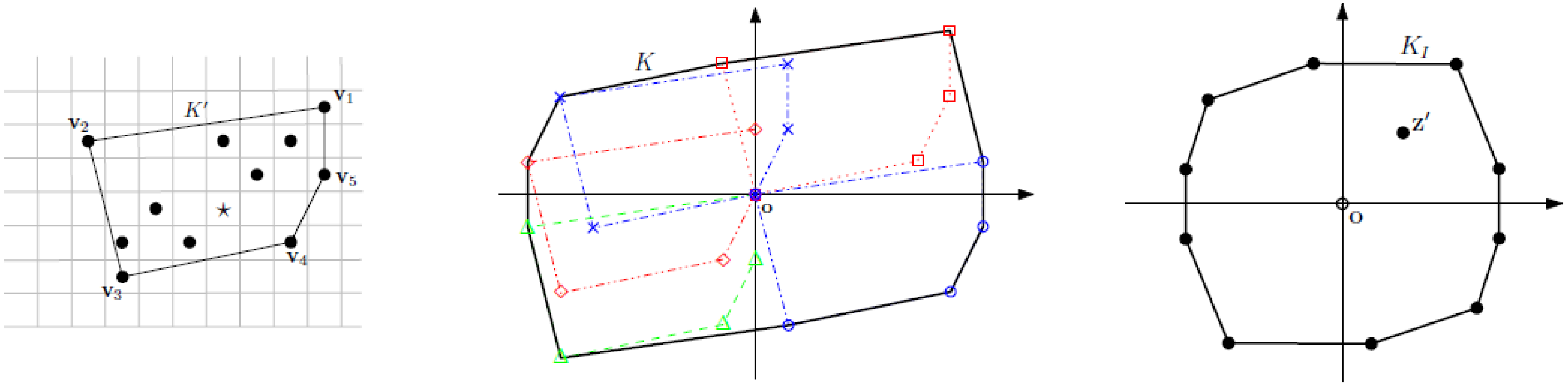
$$\frac{Pr(\mathcal{A}(\mathbf{x}_1) = \mathbf{z}_t)}{Pr(\mathcal{A}(\mathbf{x}_2) = \mathbf{z}_t)} \leq e^{\epsilon}$$



- Exponential mechanism and Laplace mechanism are not optimal

EMORY UNIVERSITY

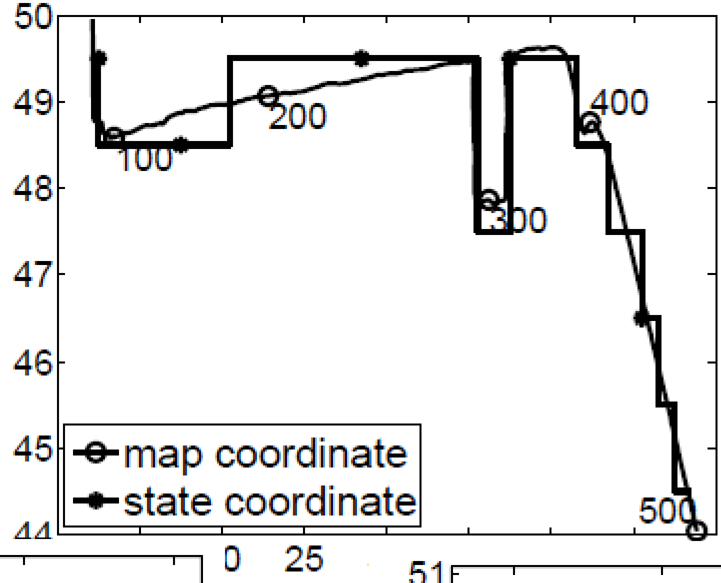
Planar Isotropic Mechanism



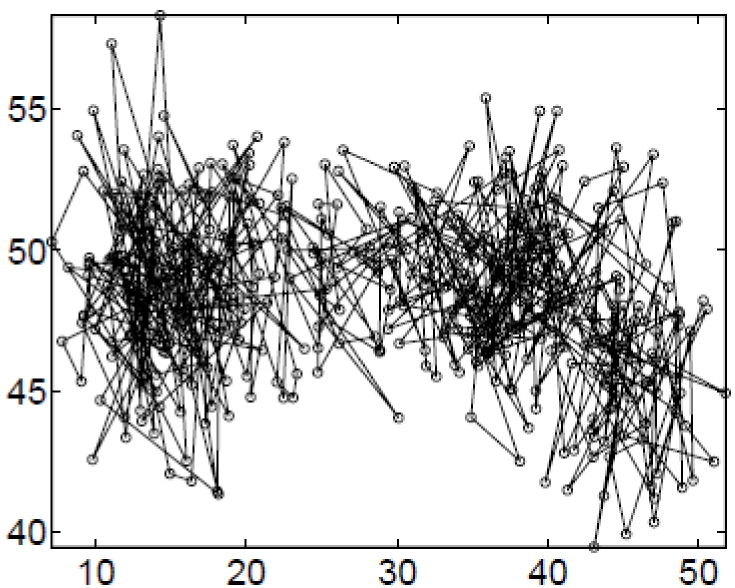
(a) (b) (c)

- Based on sensitivity hull K of δ-location set which determines the lower bound error
- An improved K-norm mechanism based on Isotropic transformation
- Achieves optimality while achieving differential privacy

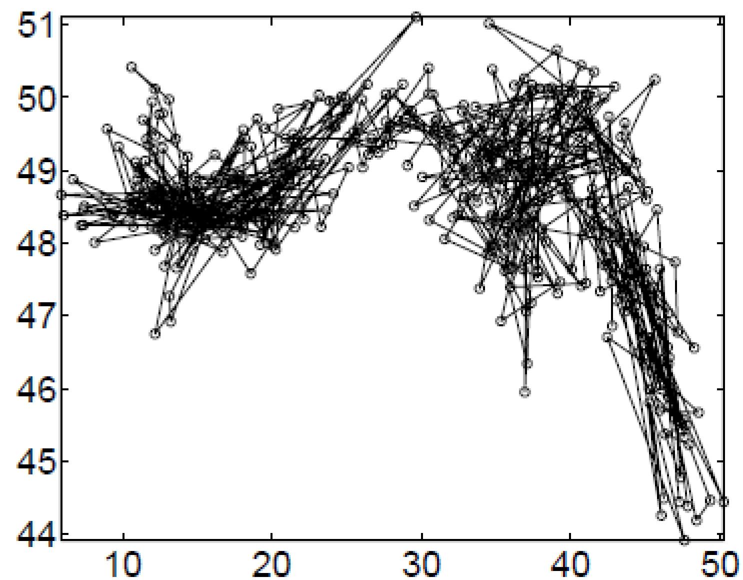
Results: Perturbed Trace Illustration



(a) Tru

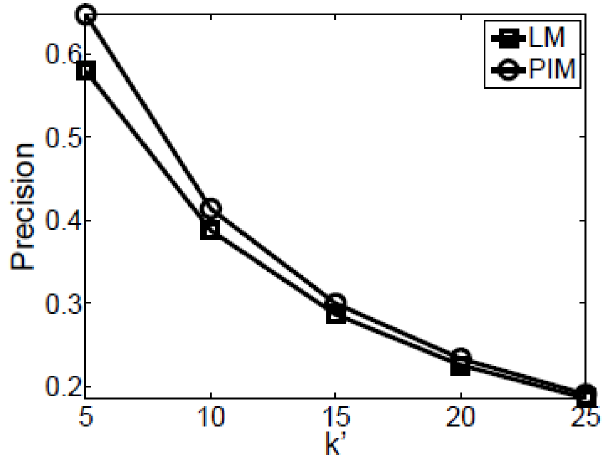


(c) LM released trace

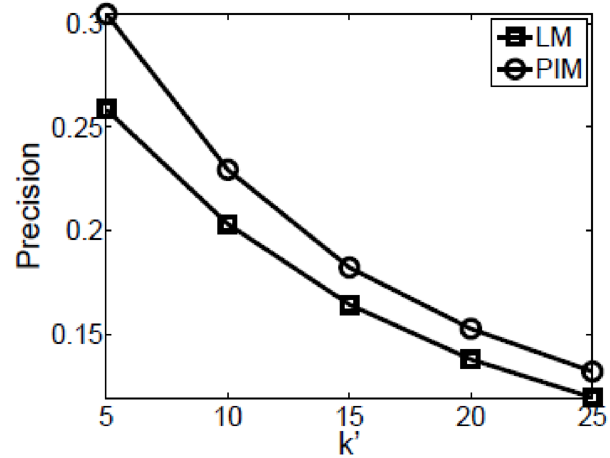


(e) PIM released trace

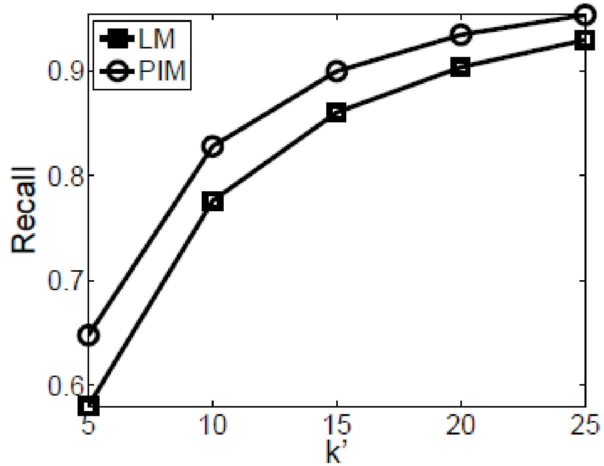
Results: k-Nearest Neighbor Queries



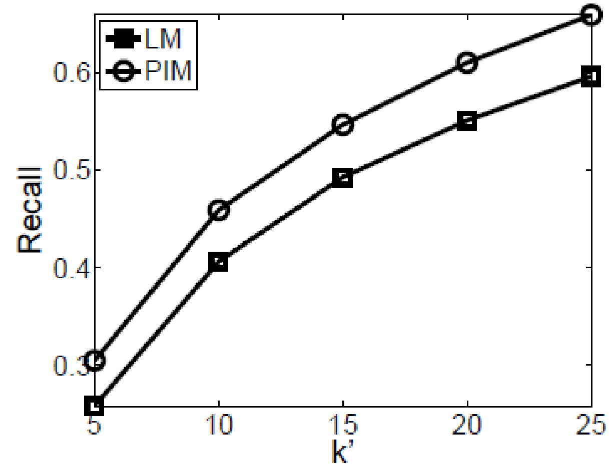
(c) Precision (Geolife)



(d) Precision (Gowalla)



(e) Recall (Geolife)



(f) Recall (Gowalla)

From Location Privacy to Spatiotemporal Privacy

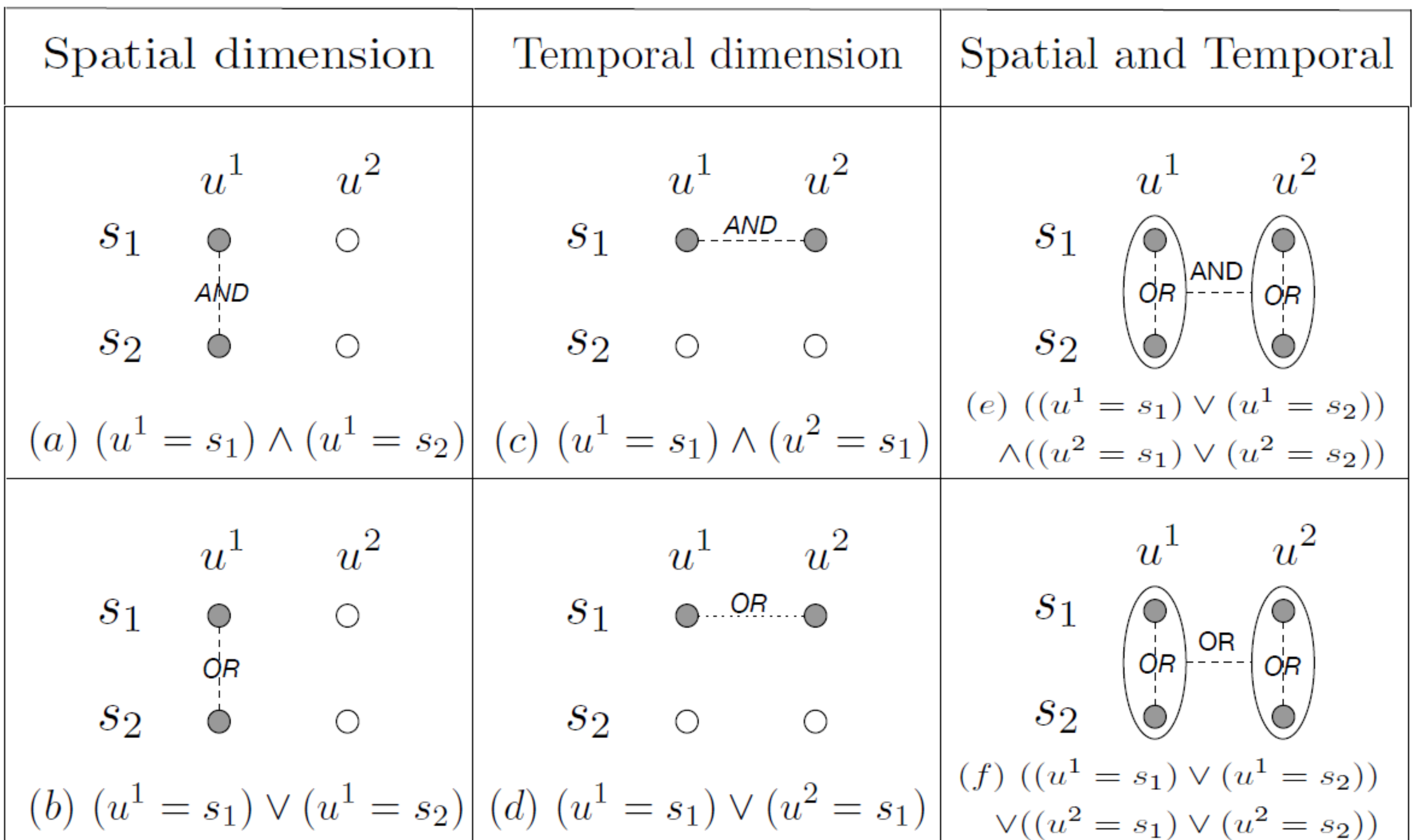
- Location privacy mechanisms protect location at a time point
- May not protect spatiotemporal activities?
  - Staying in hospital for 2 hours
  - From home to office every morning
- Need formal notions and mechanisms

Yang Cao, Yonghui Xiao, Li Xiong, Liquan Bai. PriSTE: From Location Privacy to Spatiotemporal Event Privacy (short paper). ICDE 2019

EMORY UNIVERSITY

Spatiotemporal events

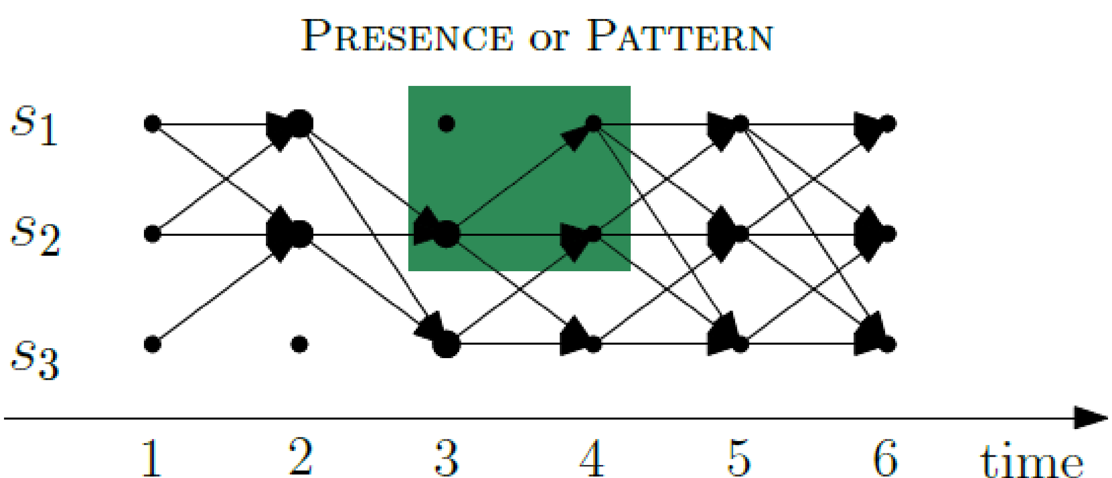
- Boolean expression for spatiotemporal event
- Location at a time point ($u^t = s_i$)



From Location Privacy to Spatiotemporal Event Privacy

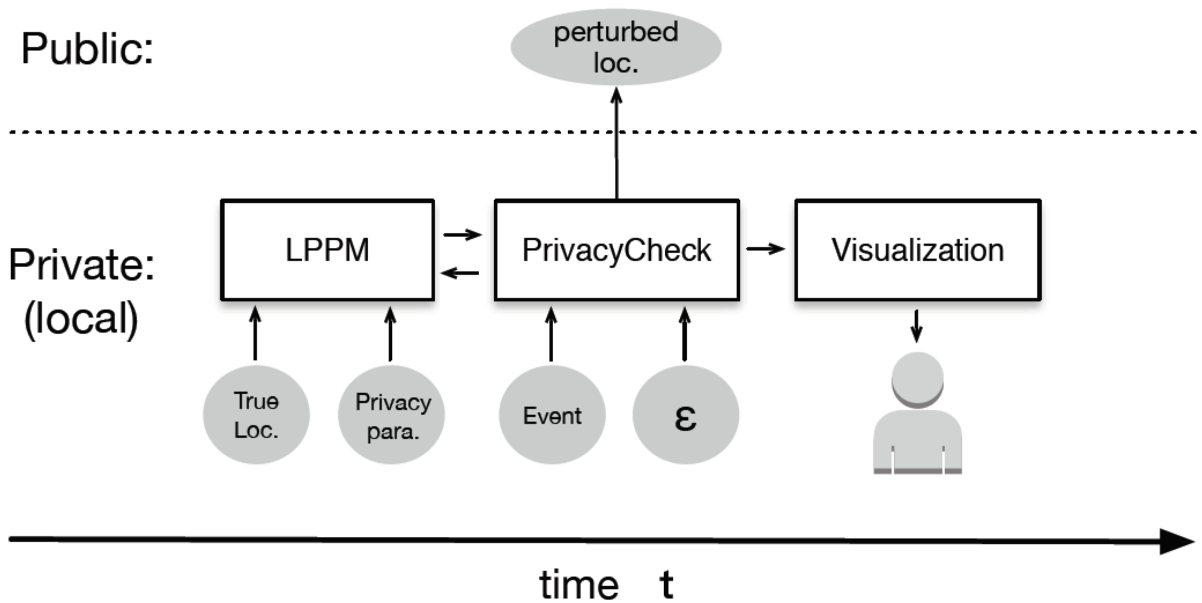
- Location privacy
  - Two locations produce “similar” distributions/observations
- Spatiotemporal event privacy
  - A true event and a negative event produce “similar” location traces

$$Pr(o_1, o_2, \cdots, o_t | \textsc{Event}) \leq e^{\epsilon} Pr(o_1, o_2, \cdots, o_t | \neg\textsc{Event})$$



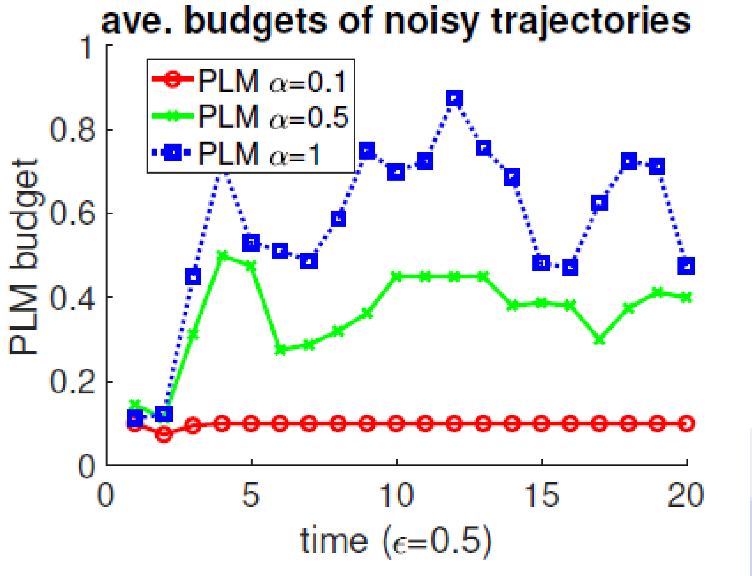
Spatiotemporal Privacy Framework

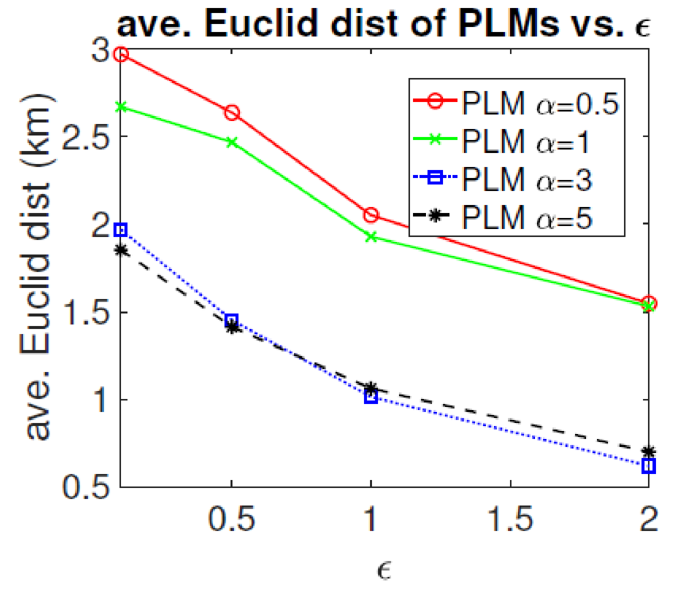
- **LPPM**: Existing location privacy mechanism, e.g. Planar Laplace Mechanism for geo-indistinguishibility
- **PrivacyCheck**: check spatiotemporal event privacy and calibrate privacy budget



Results

- Strong LPPM may satisfy spatiotemporal privacy already
- Weak LPPM need to reduce privacy budget significantly (less utility) to achieve same level of spatiotemporal privacy
- Stronger spatiotemporal privacy, less utility of the locations



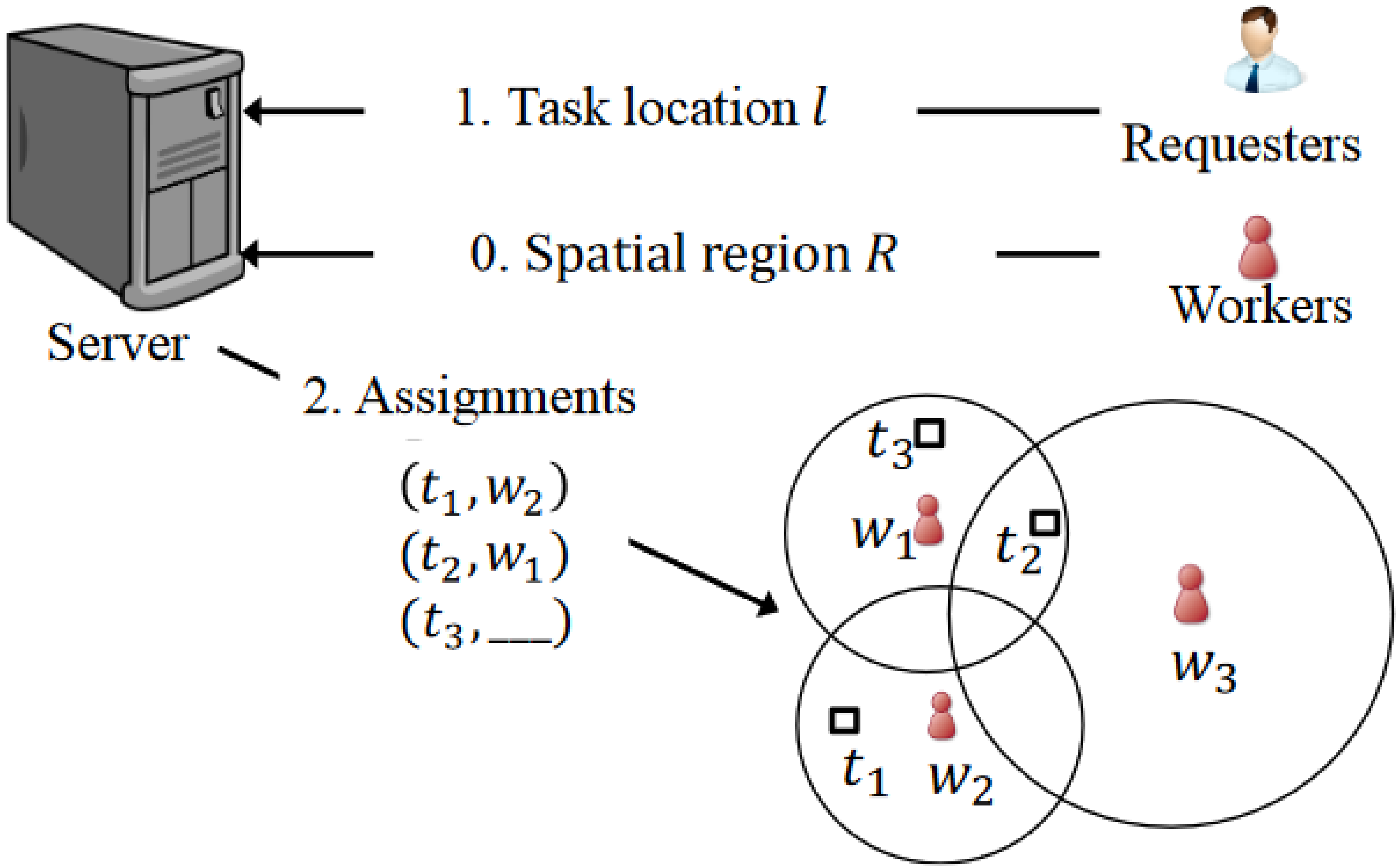


EMORY UNIVERSITY

Enabling Mobile Apps and Analytics with Local Differential Privacy

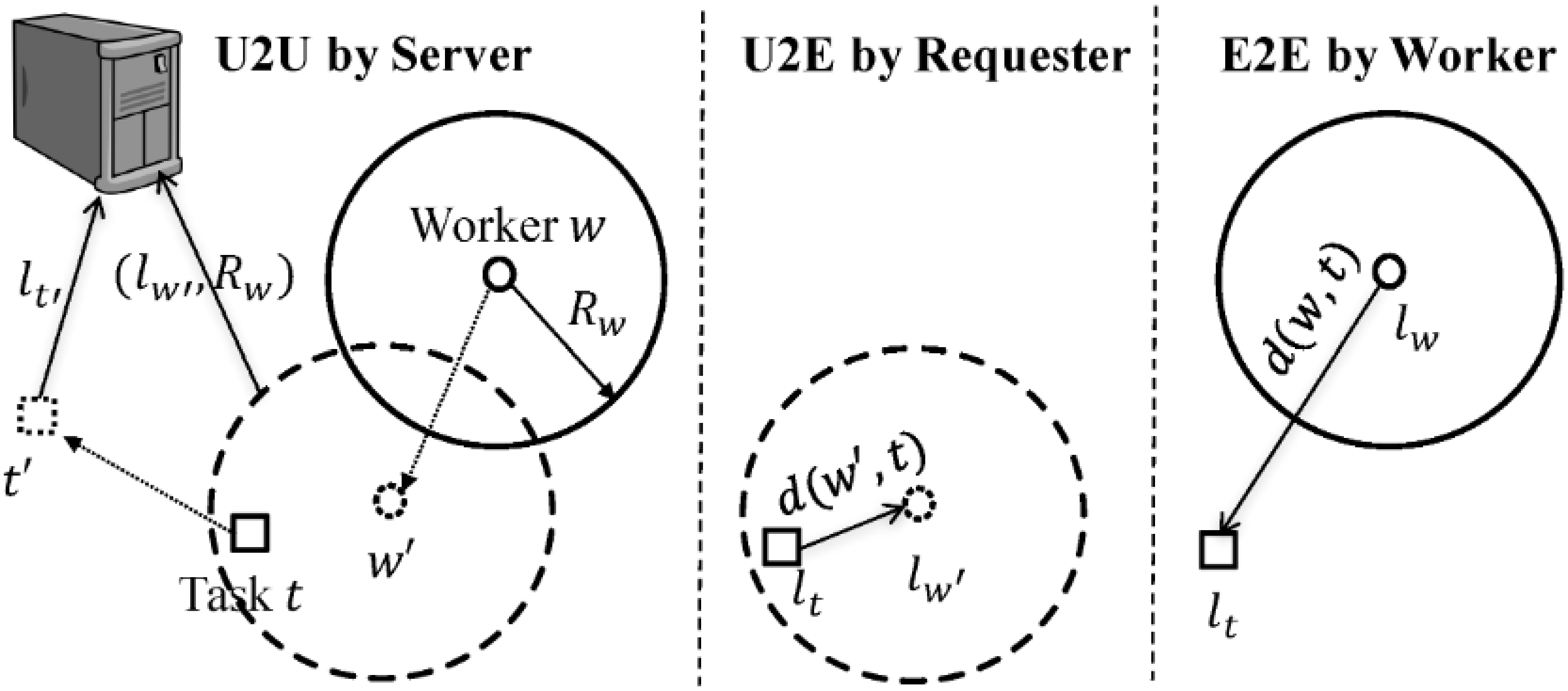
- Background
  - Local differential privacy
  - Geo-indistinguishability (local d-privacy)
- Extended privacy notions
  - Protecting dynamic locations (CCS15, VLDB17 demo)
  - Protecting spatiotemporal events (ICDE19)
- New mobile applications
  - Spatial crowdsourcing with geo-indistinguishability (ICDE18)
- New mechanisms
  - Supporting both analytics and mobile applications (CNS19)

ONLINE TASK ASSIGNMENT IN SPATIAL CROWDSOURCING



Privacy preserving online task assignment in spatial crowdsourcing

- Both requester and worker locations are perturbed using geo-indistinguishability
- Three-stage framework for task assignment using uncertain locations



Hien To, Cyrus Shahabi, Li Xiong. Privacy-Preserving Online Task Assignment in Spatial Crowdsourcing with Untrusted Server. ICDE 2018

Enabling Mobile Apps and Analytics with Local Differential Privacy

- Background
  - Local differential privacy
  - Geo-indistinguishability (local d-privacy)
- Extended privacy notions
  - Protecting dynamic locations (CCS15, VLDB17 demo)
  - Protecting spatiotemporal events (ICDE19)
- New mobile applications
  - Spatial crowdsourcing with geo-indistinguishability (ICDE18)
- New mechanisms
  - Supporting both analytics and mobile applications (CNS19)

EMORY UNIVERSITY

Supporting both range queries and frequency estimation

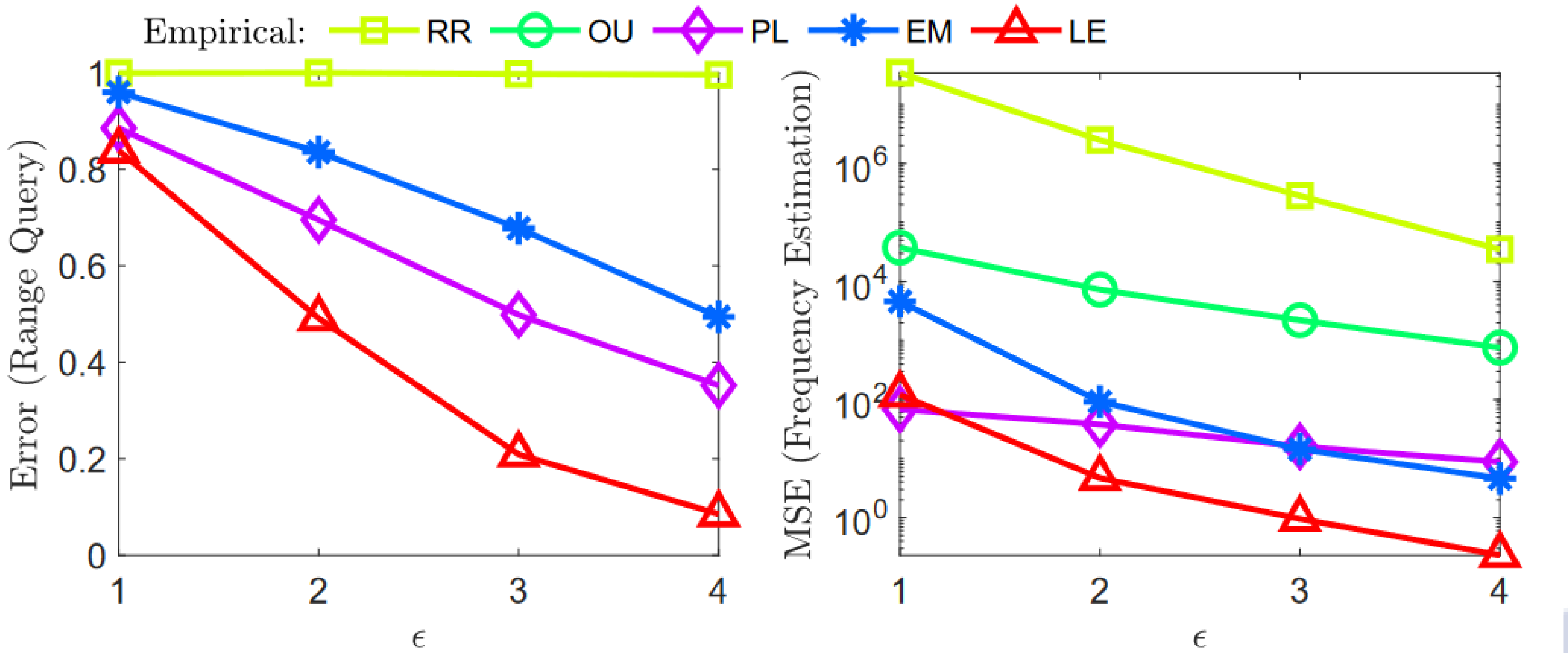
- Existing
  - Local differential privacy with randomized response – frequency estimation
  - Geo-indistinguishability (local d-privacy) with planar Laplace mechanism – range queries
- Goal
  - Optimize for both frequency estimation and range queries while ensuring local d-privacy
- Basic idea
  - Assign different perturbation probabilities for different input/output pairs in a way related to the distance

X. Gu, M. Li, Y. Cao and L. Xiong, Privacy-Preserving Range Queries and Frequency Estimation with Geo-indistinguishability. IEEE Conference on Communications and Network Security (CNS), 2019

EMORY UNIVERSITY

Results: Comparison

Gowalla dataset



RR: Randomized Response
OU: Optimized with Unary Encoding
PL: Planar Laplace mechanism
EM: Exponential mechanism
LE: Linear equation mechanism

EMORY UNIVERSITY

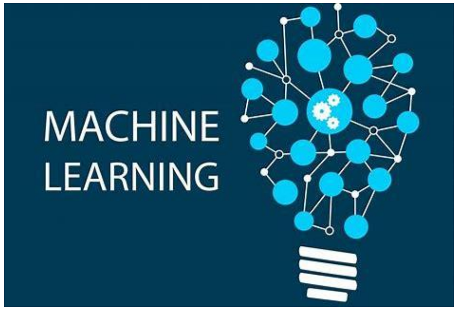
Enabling Mobile Apps and Analytics with Local Differential Privacy

- Extended privacy notions
  - Protecting dynamic locations (CCS15, VLDB17 demo)
  - Protecting spatiotemporal events (ICDE19)
- New mobile applications
  - Spatial crowdsourcing with geo-indistinguishability (ICDE18)
- New mechanisms
  - Supporting both analytics and mobile applications (CNS19)
- Open challenges
  - Privacy/utility tradeoff
  - User empowerment

EMORY UNIVERSITY

Assured Information Management and Sharing (AIMS)

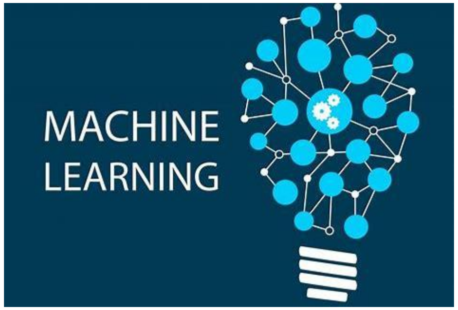
+



+

Assured Information Management and Sharing (AIMS)

+



+

http://www.cs.emory.edu/site/aims

EMORY
UNIVERSITY