A Closer Look: Evaluating Location Privacy Empirically

Liyue Fan, Assistant Professor in Computer Science

UNC Charlotte

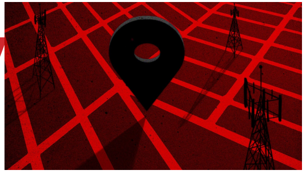
PEPR 2022

Outline

- Location privacy: state of research
- Challenges of adoption
- Local, online privacy methods
- Evaluation: methods and data
- Results *new*
- Discussion and take-aways

Fan and Gote. "A Closer Look: Evaluating Location Privacy Empirically". In SIGSPATIAL'21.

Importance of Location Privacy



- Location data enables numerous applications
  - recommendations [Levandoski et al. 2012]
  - mental health research [Canzian and Musolesi 2015, Palmius et al. 2016]

- Location data is sensitive
  - home/work location
  - visit to a hospital, political or religious event

- **Location privacy solutions are needed.**

17 MAY 18

Tracking Firm LocationSmart Leaked Location Data for Customers of All Major U.S. Mobile Carriers Without Consent in Real Time Via Its Web Site

Locat… real-time data about the

precis… rmation to anyone via a

KrebsonSecurity 2018

FTC Brings First Case Against Developers of “Stalking” Apps

October 22, 2019

Settlement resolves charges that Retina-X’s products created security vulnerability and violated consumers’ privacy

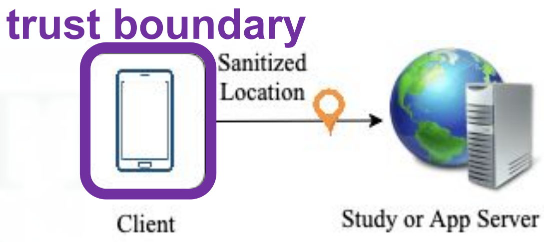
FOR RELEASE

ftc.gov

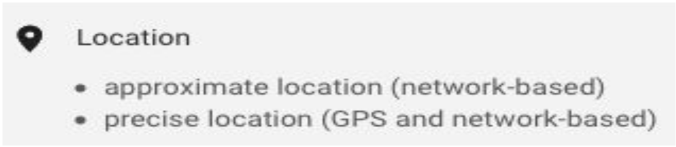
Location Privacy Research

- Numerous location privacy methods proposed in the last two decades
  - 60+ studied in [Primault et al. 2019]

- ***Our focus: local & online***
  - users have a sense of control
  - data is available immediately

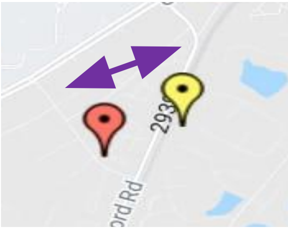


- Promise for practical deployment
  - Analogous to the LDP model for non-location data
  - Android users specify location sharing preferences for apps
  - Effort to open-source local online privacy methods, e.g., Geopriv4j [Fan and Gunja 2020]



Goals of this work

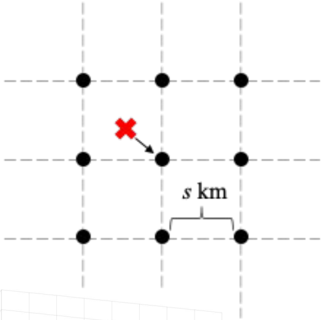
Adoption of Location Privacy



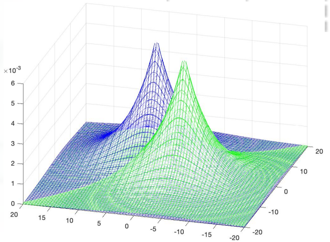
- **Challenge 1: understand the impact of location privacy on usefulness**
  - Prior studies evaluate *simple* measures
  - Not clear how location privacy may affect *applications*
- **Challenge 2: understand the empirical privacy protection**
  - Privacy models of existing methods are *not comparable*
  - Not clear how current methods mitigate *practical attacks*
- **Challenge 3: understand computational overheads**
  - Important for deployment but under-studied

Local, Online Privacy Methods

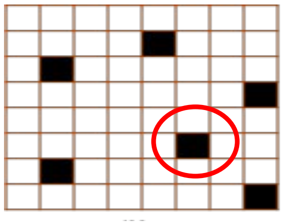
- Generalization-based: report approximate data
  - Rounding [Krumm 2007, Micinski et al. 2013]
  - Spatial Cloaking [Krumm 2007]



- Perturbation-based: “add” noise to data
  - Noise [Krumm 2007]
  - Various-size Hilbert Curve [Pingley et al. 2009]
  - Geo-indistinguishability (Laplace) [Andrés et al. 2013]



- Dummy-based: hide among dummies
  - SpotME [Quercia et al. 2011]
  - Moving in the Neighborhood [Kido et al. 2005]



Data & Method

- Real GPS traces

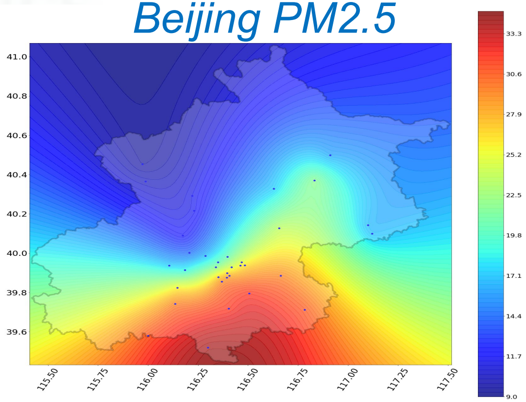
Table 1: Dataset Summary

Dataset	#Users	Frequency	Resolution	Avg. # Traj's	Avg. # Loc's
GeoLife[25]	182	1 to 5 seconds	182×182	54	15640
RioBuses[6]	14149	every minute	170×170	9	2661

- Preprocessing
  - Spatial discretization: 2D grid cells, ~300m x 300m each
  - Temporal: subsample every 5 minutes

- Applications (*new*)
  - Co-location detection
  - Air pollution exposure

Beijing PM2.5



Utility Measures

- Record-level errors vs. trace-level mobility pattern errors

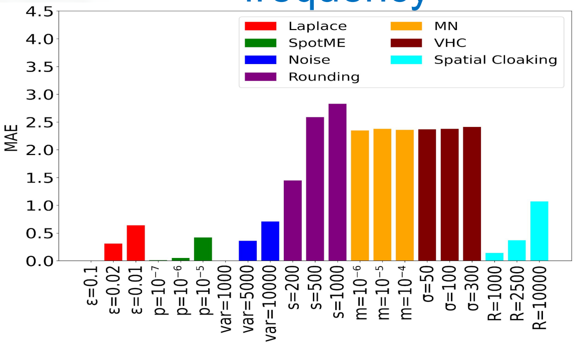
[Canzian and Musolesi 2015]

- Frequency & 2D range queries

Table 3: Laplace Utility Experiment - GeoLife

Utility/Params	ϵ					
	0.001	**0.01**	**0.02**	**0.04**	**0.05**	**0.1**
Hamming	0.74	0.41	0.20	0.03	0.01	0.00
Haversine (in m)	1494.96	121.57	46.52	7.34	2.83	0.02
Tot Dist (in %)	99.18	91.16	73.06	18.14	6.43	0.00
Max Dist (in %)	98.25	89.98	72.27	17.88	6.15	0.00
Std Dev Displacement (in %)	98.74	85.26	58.36	12.85	4.47	0.00
Max Dist Home (in %)	69.94	26.41	16.68	2.97	0.00	0.00
Rad Gyration (in %)	98.02	89.84	72.40	17.92	6.18	0.00
# Diff Places (in %)	96.87	90.66	72.94	18.16	6.15	0.00
# Significant Places (in %)	14.97	22.75	12.57	3.59	1.20	0.00
Avg Mobility Error (in %)	82.28	70.87	54.04	13.07	4.37	0.00

frequency



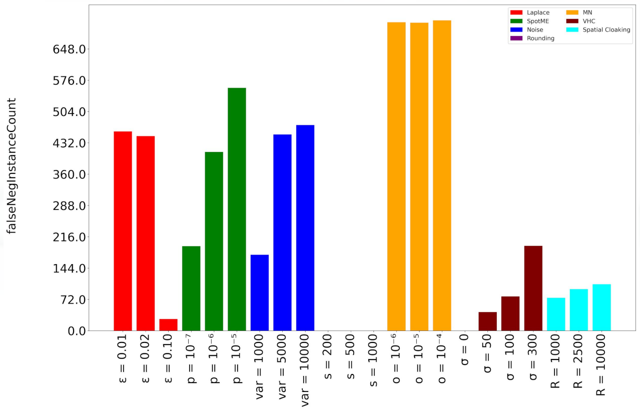
Record-level utility often, but not always, aligns with trace-level utility

Utility for Applications

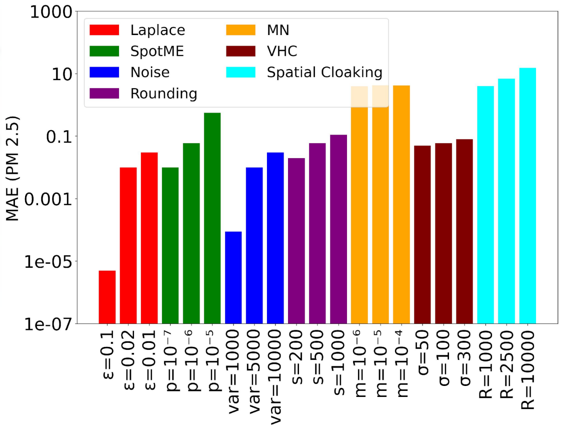
new

- Co-location detection
- Air pollution exposure

GeoLife - False negative user pairs



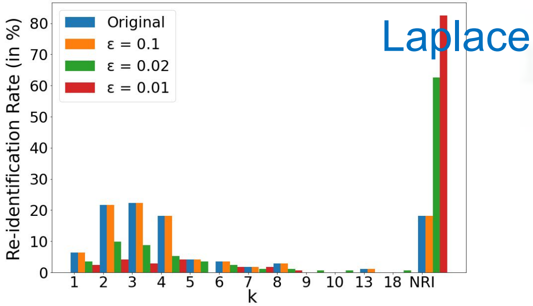
GeoLife - PM2.5



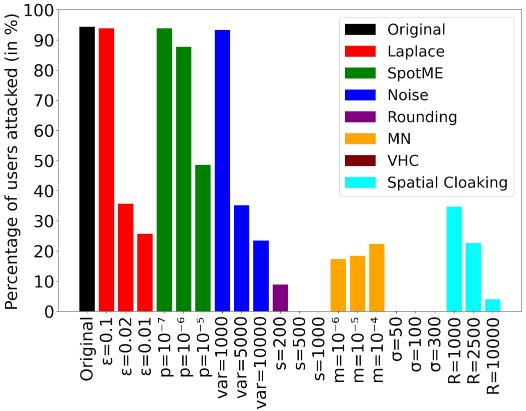
- *Choosing privacy methods & params is important for utility.*

Empirical Privacy Measures

- Re-identification attack
  - Knowing any ***k*** locations of the target, how likely is the target **uniquely** identified?
  - We find the smallest k for each user



- Inference attack
  - Knowing **all but one** locations of the target, how likely to infer the **last location**?

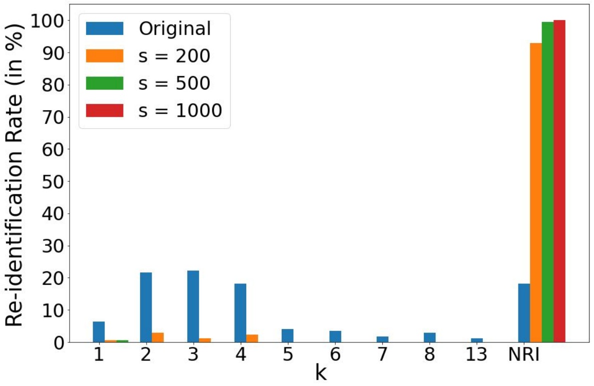


Both DP and traditional methods provide protection against attacks.

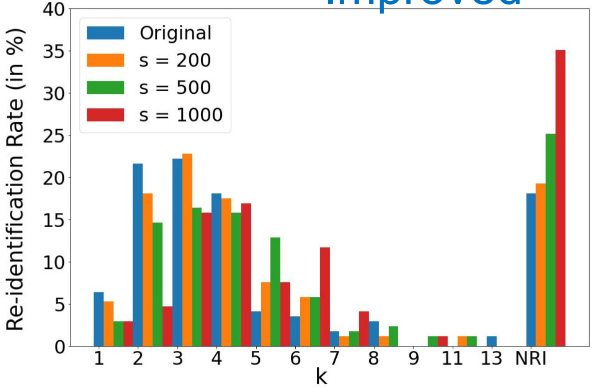
Improved Attacks

- An adversary knows the privacy method & param value
- *Rounding* results

Basic



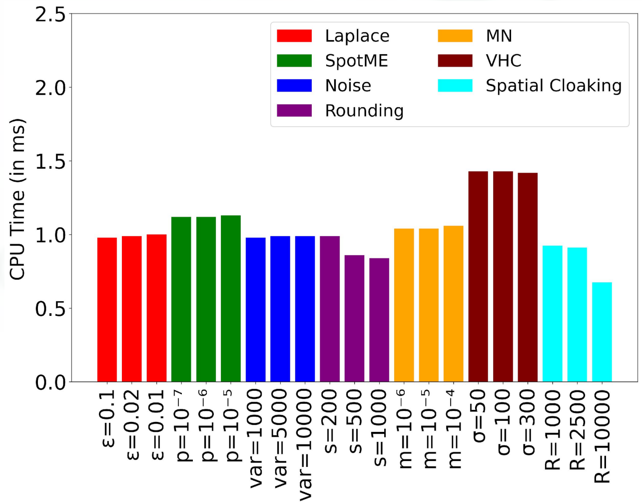
Improved



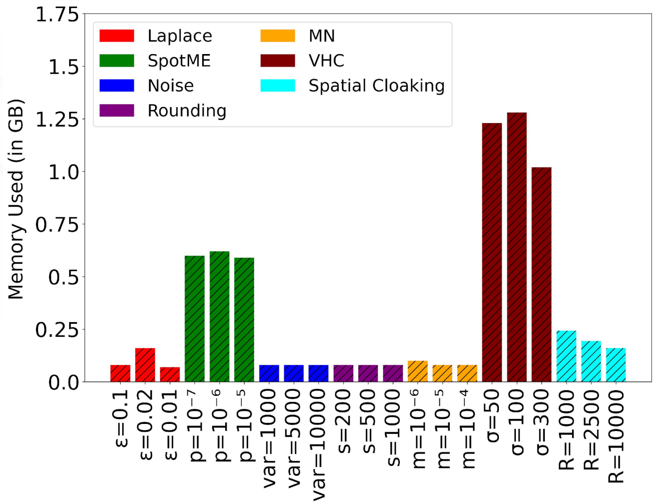
Traditional, deterministic methods may fail to protect privacy in improved attack.

Overheads

- CPU time to sanitize each location



- Peak memory requirement



All methods are very efficient in CPU time.

Conclusions and Discussion

- This study enables app developers and researchers to comparatively evaluate existing location privacy methods
- All methods are open-sourced in Java

fan-group / geopriv4j

- Generic utility often but not always aligns with task-based utility
- Basic attacks: both differential privacy-based and traditional methods provide protection
- Improved attacks: deterministic methods may fail to provide adequate protection
- Choosing the right methods and params is important
- Many studied methods have low CPU and memory requirements

Acknowledgements

- Students
  - Grad alum: Sriram, Ishan, Julius
  - Undergrads: Ethan, Ashley

- Questions? Contact Liyue liyue.fan@uncc.edu

CNS-1951430