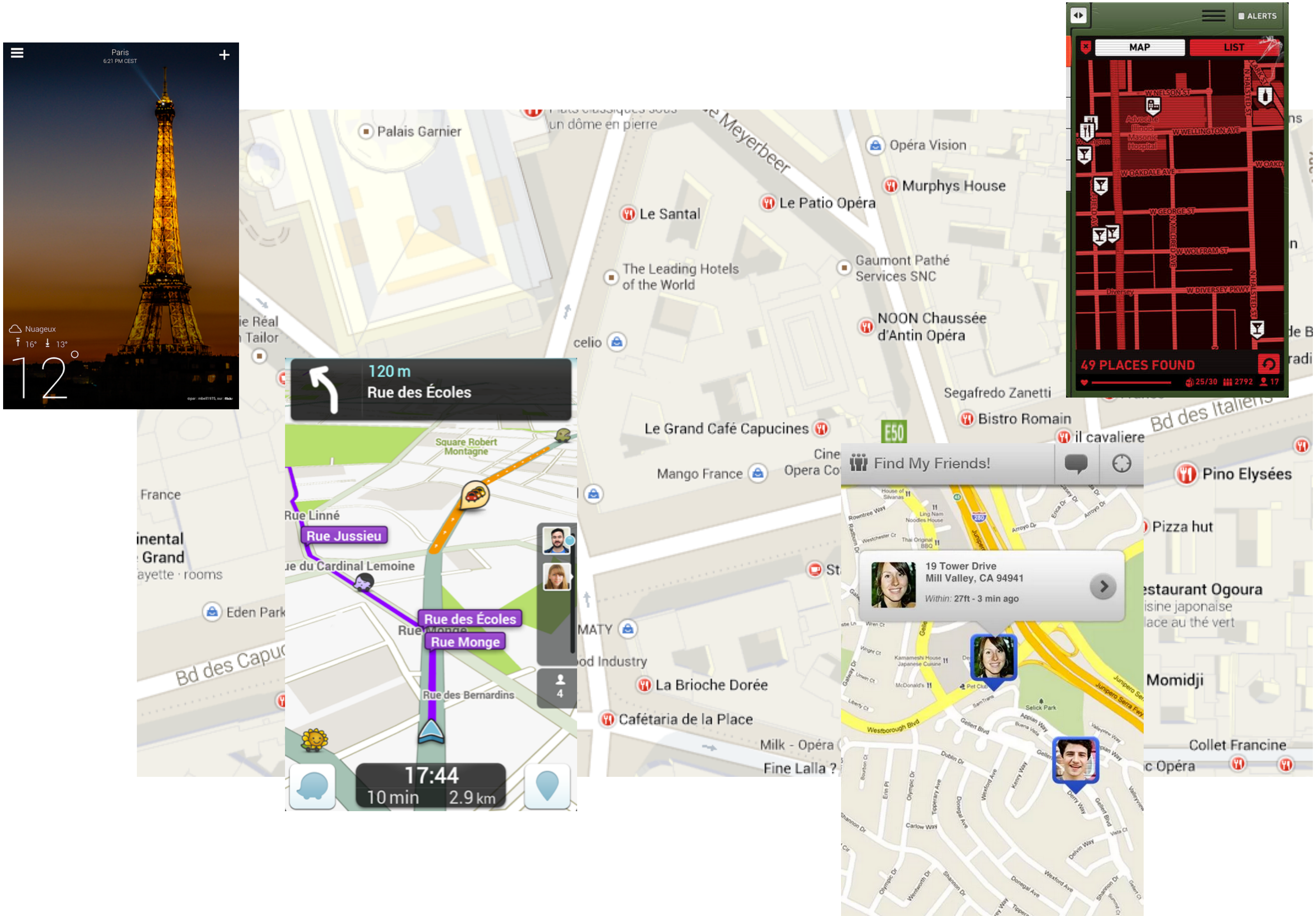
Differentially Private Location Privacy in Practice

Vincent Primault
Sonia Ben Mokhtar
Cédric Lauradoux
Lionel Brunie

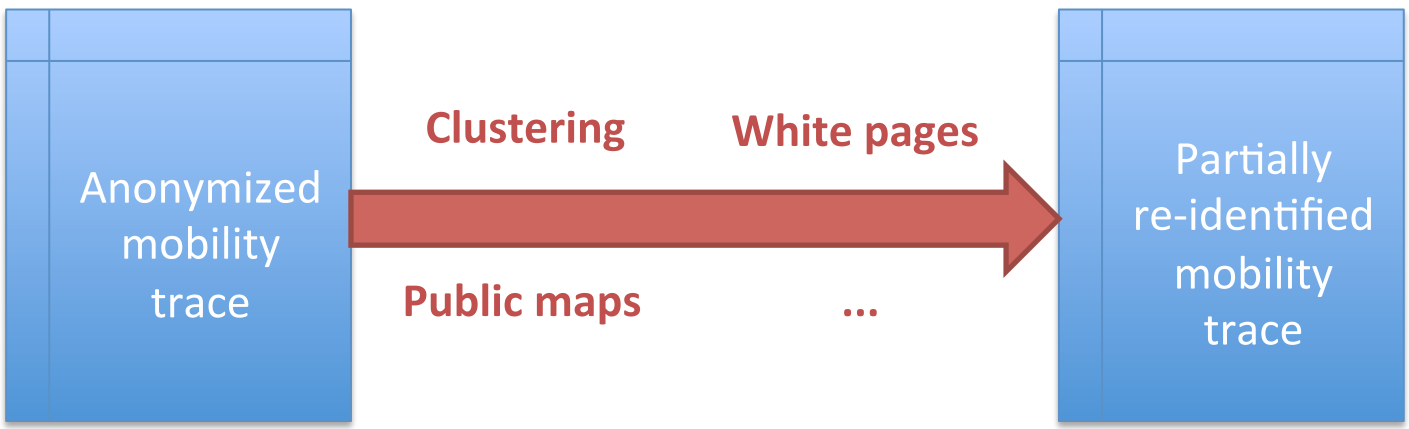
May 17th 2014

Location-based services




PLEASE ROB ME
Raising awareness
about over-sharing
Check out our guest blog post on the CDT website.

Location privacy threats



Only 4 points are sufficient to uniquely identify you! [1]

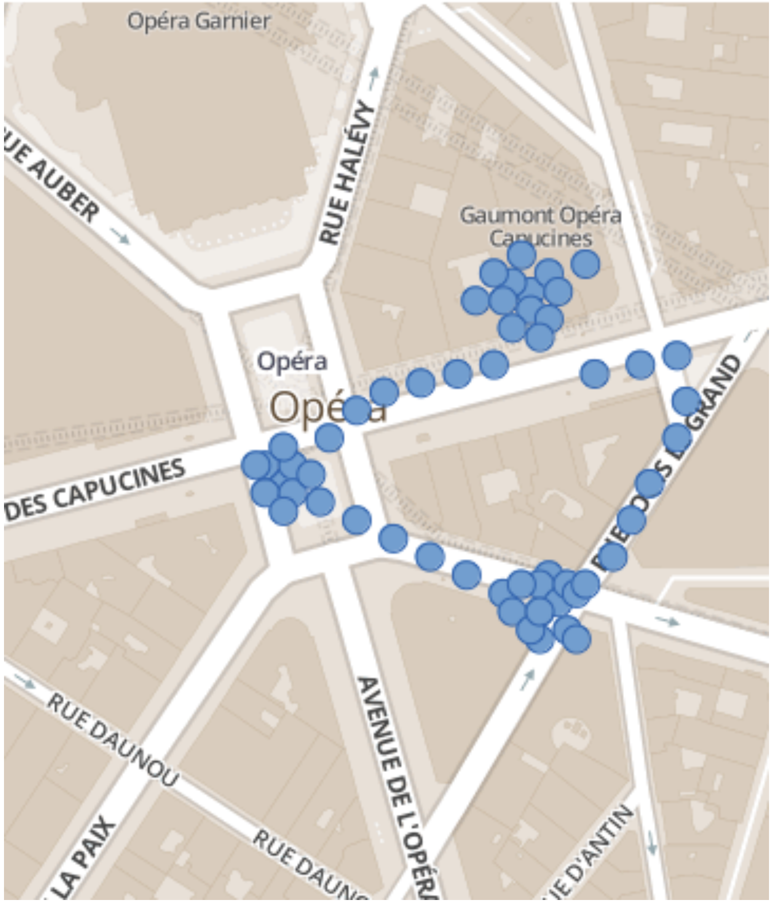
[1] De Montjoye et al. **Unique in the Crowd: The privacy bounds of human mobility**. *Scientific reports, 2013.*
[2] Golle et al. **On the Anonymity of Home/Work Location Pairs.** *Pervasive'09.*

Can a protection mechanism efficiently protect points of interest of a user?

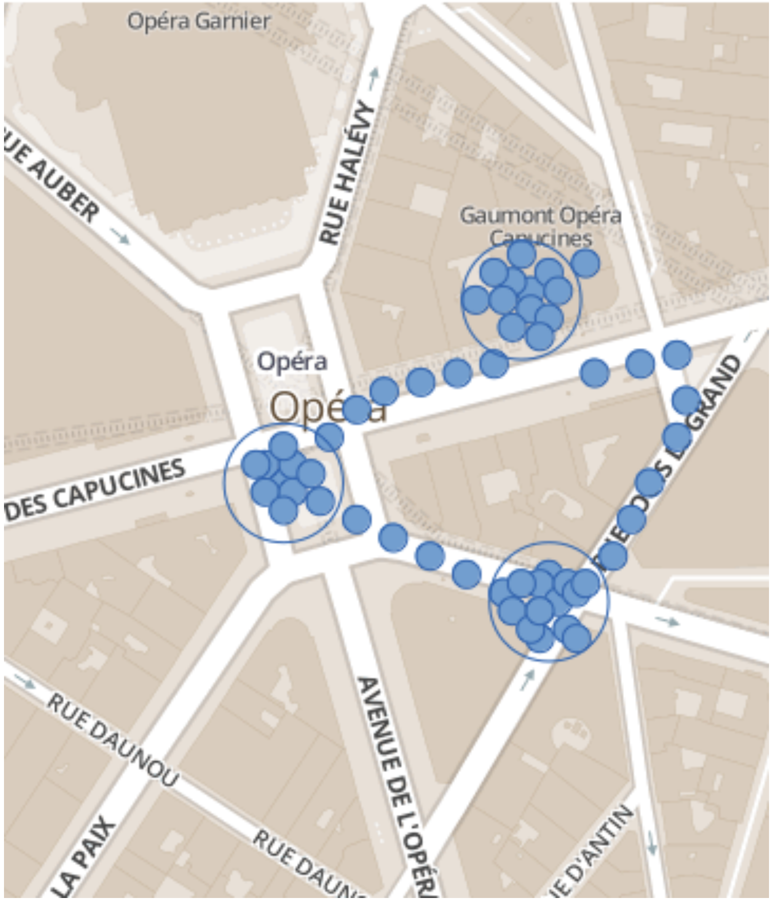
Outline

- Introduction
- **About points of interest**
- Protection mechanisms
- Experimental settings
- Evaluation metrics & results
- Sum-up

A mobility trace



Areas of interest



Points of interest

Outline

- Introduction
- About points of interest
- **Protection mechanisms**
- Experimental settings
- Evaluation metrics & results
- Sum-up

Location-privacy protection mechanisms

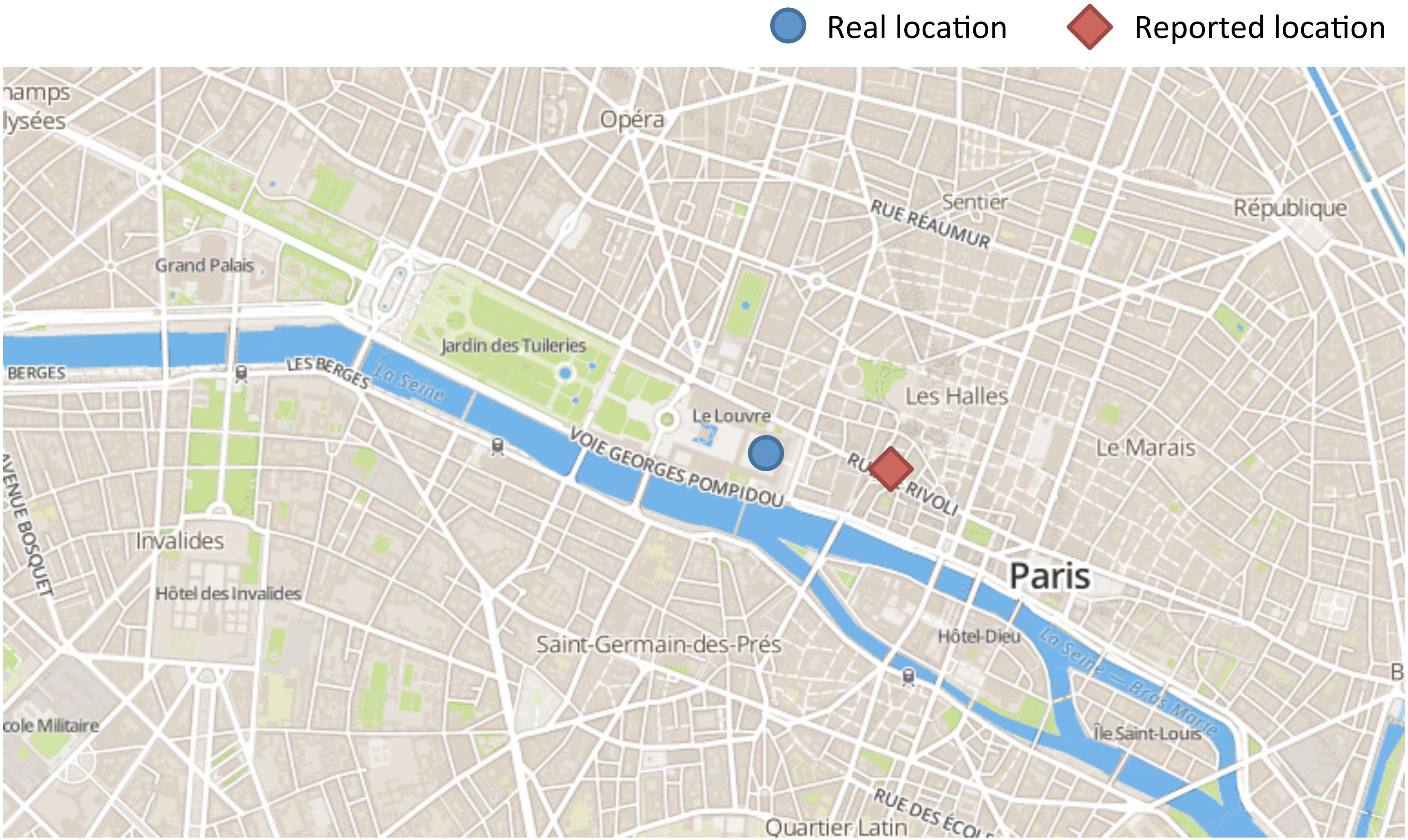
Pseudonymity
Mix-zones

Spatial cloaking
k-anonymity

Noise-based solutions

Cryptographic protocols

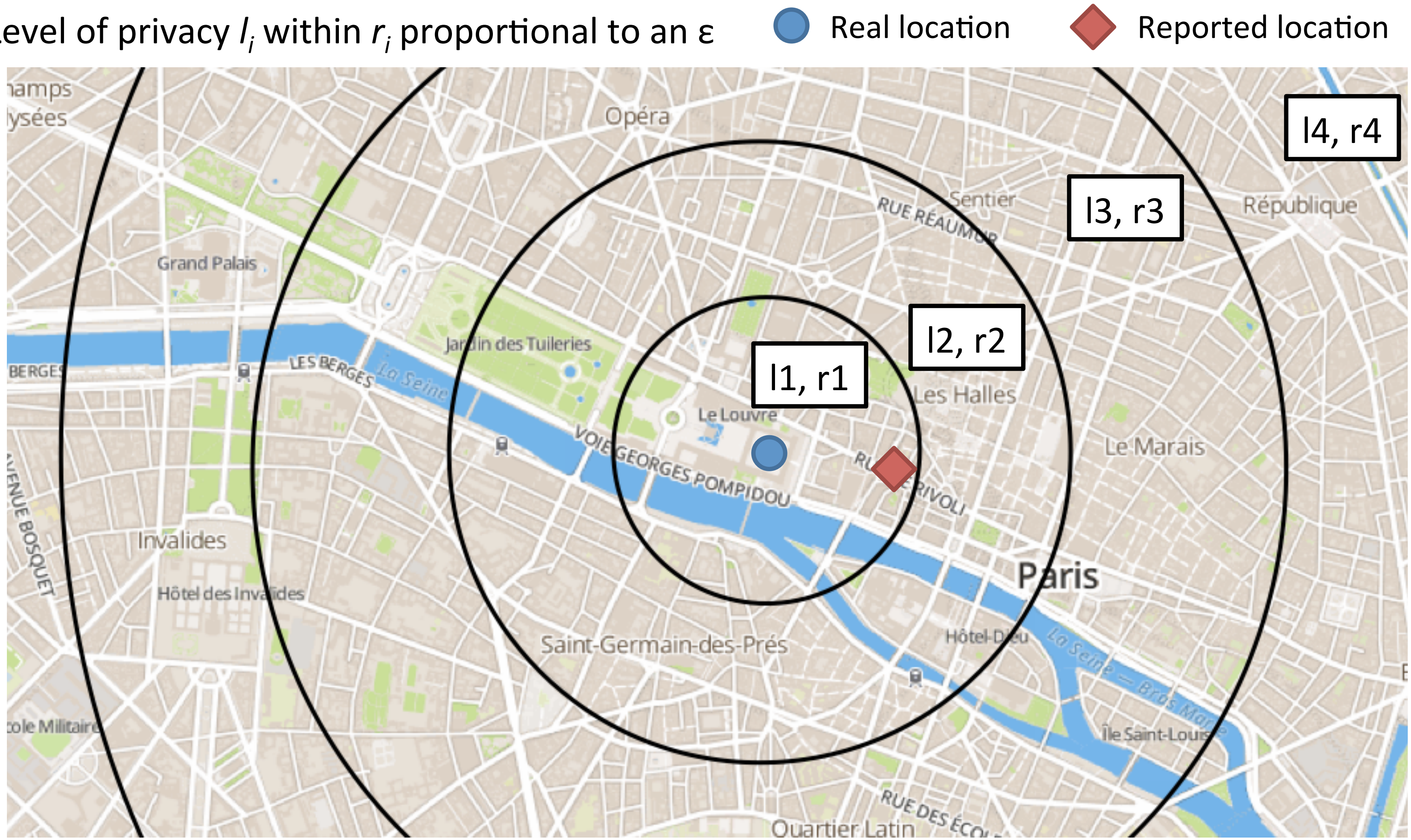
Geo-indistinguishability



[3] Andrés et al. **Geo-indistinguishability: Differential privacy for Location-based Systems**. *CCS'13.*

Geo-indistinguishability

Level of privacy l_i within r_i proportional to an ε



[3] Andrés et al. **Geo-indistinguishability: Differential privacy for Location-based Systems**. *CCS'13.*

Outline

- Introduction
- About points of interest
- Protection mechanisms
- **Experimental settings**
- Evaluation metrics & results
- Sum-up

Two different data sets

San Francisco cabs
In the SF Bay Area
1 month in 2009
536 taxis
11 millions points

Geolife
Around Beijing
4 years (2007-2011)
182 users
25 millions points

Reduced Geolife
Around Beijing
1 continuous month
61 users
5 millions points

POIs extraction algorithm

Mobility trace — Time-ordered list of locations

1 hour

?

Extract stays [4] — Centroids of areas where a user has spent at least ***minTime*** within a ***maxDistance*** radius

Group nearby stays together [5] — Stays within ¾ *maxDistance* where a user passed through at least ***minPts*** times

2 times

POIs — A set of important places for a user

[4] Hariharan et al. **Project Lachesis: parsing and modeling location histories.** *GIScience'04.*
[5] Zhou et al. **Discovering Personal Gazetteers: An Interactive Clustering Approach**. *GIS'04.*

Playing with distance threshold

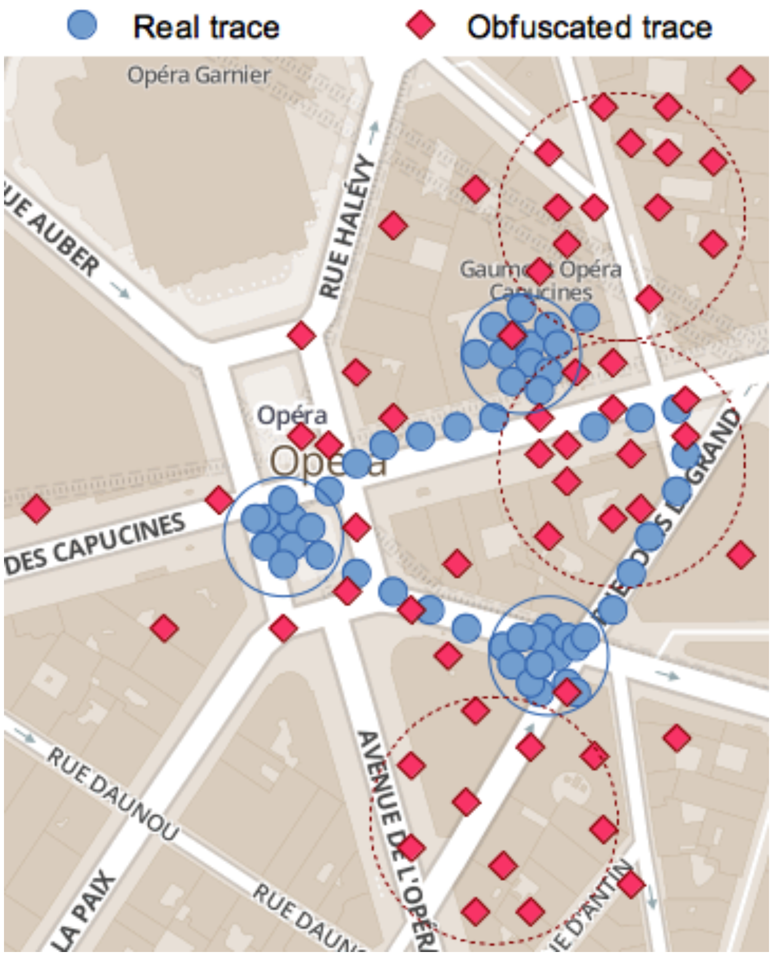
	SF cabs	Geolife
Unobfuscated	***250 m***	***250 m***
Weak privacy	700 m	600 m
Medium privacy	1000 m	1200 m
Strong privacy	2000 m	2500 m

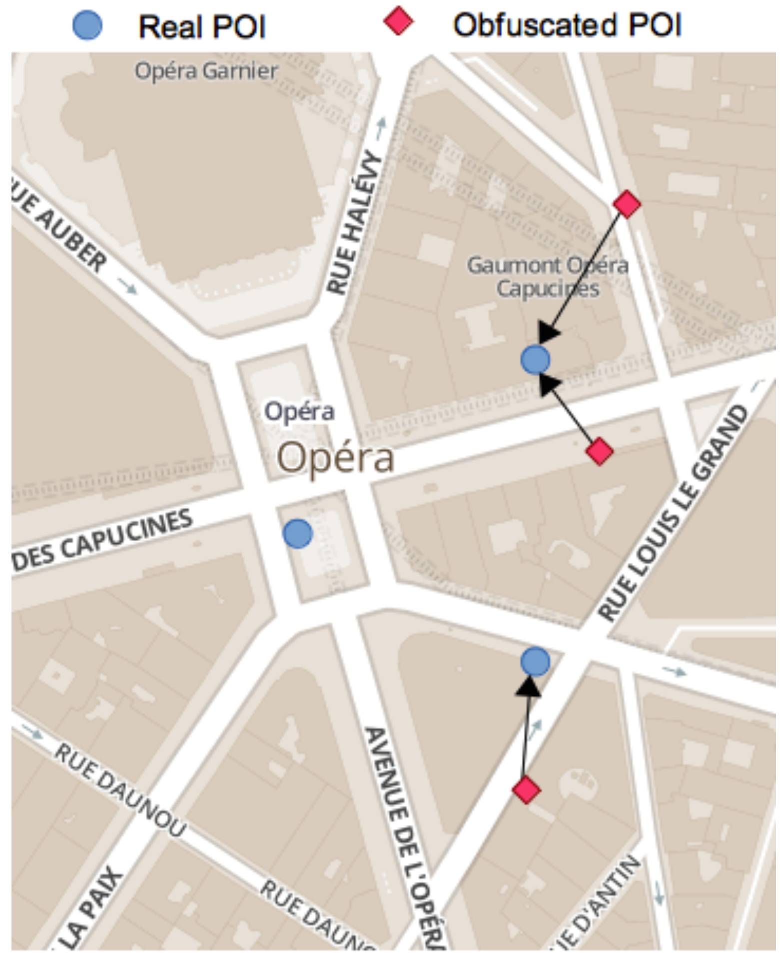
We must greatly increase the ***maxDistance*** threshold at highest privacy levels in order to retrieve an interesting number of POIs.

Outline

- Introduction
- About points of interest
- Protection mechanisms
- Experimental settings
- **Evaluation metrics & results**
- Sum-up

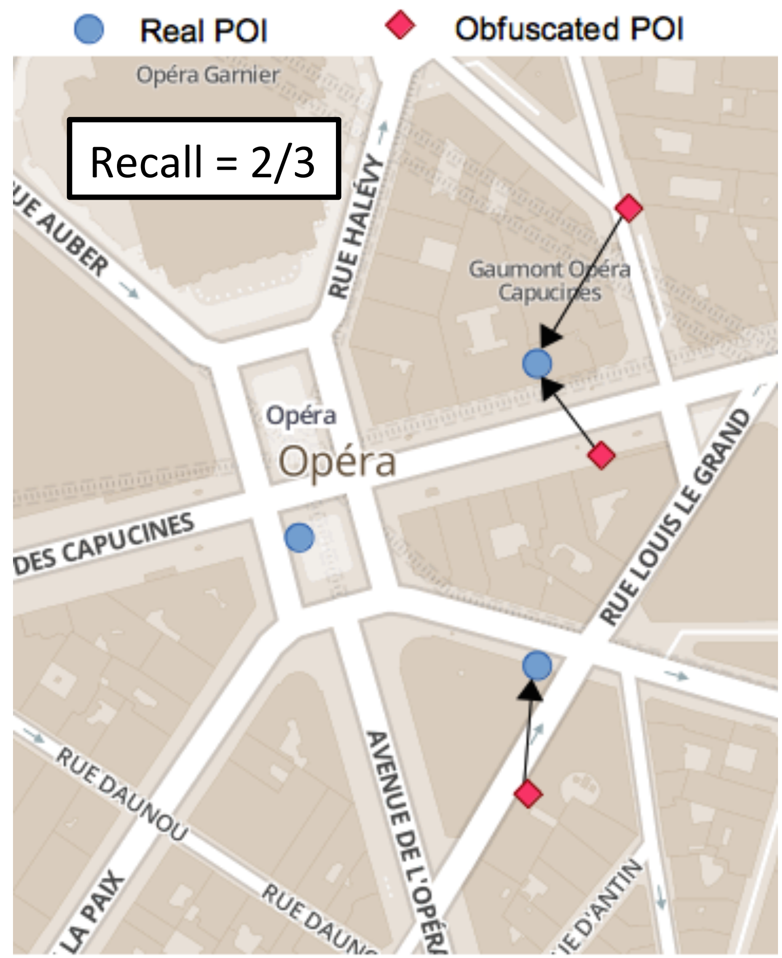
Measuring privacy





Recall rate

Recall rate is the proportion of real POIs successfully retrieved.



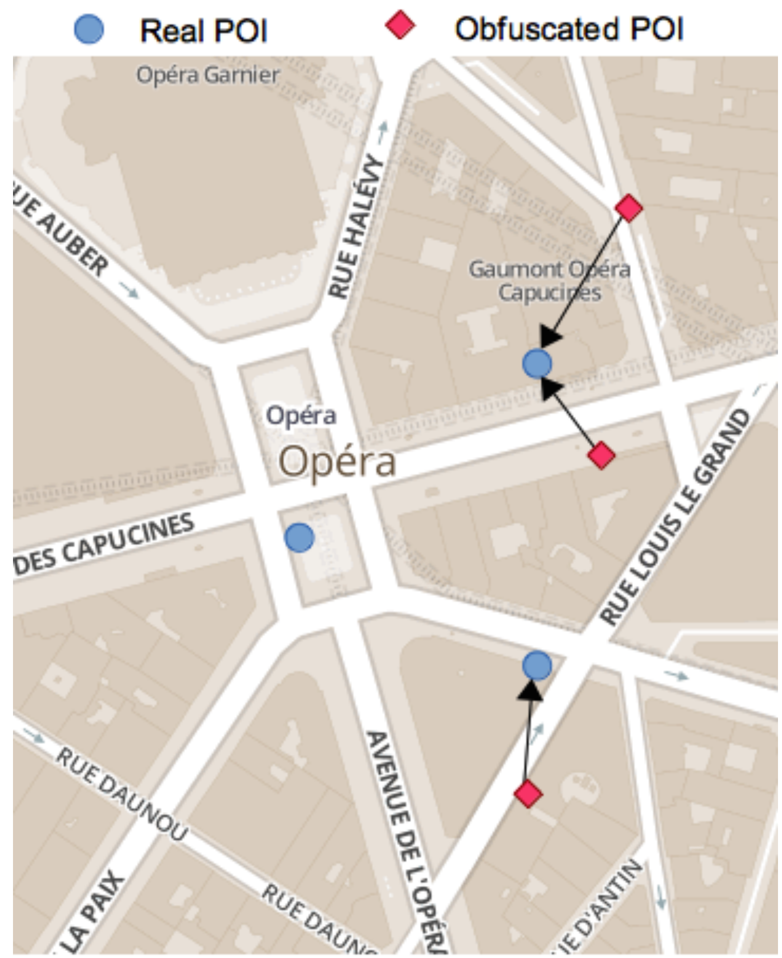
Recall rate

	SF cabs	Geolife
Weak privacy	73 %	72 %
Medium privacy	72 %	71 %
Strong privacy	71 %	61 %

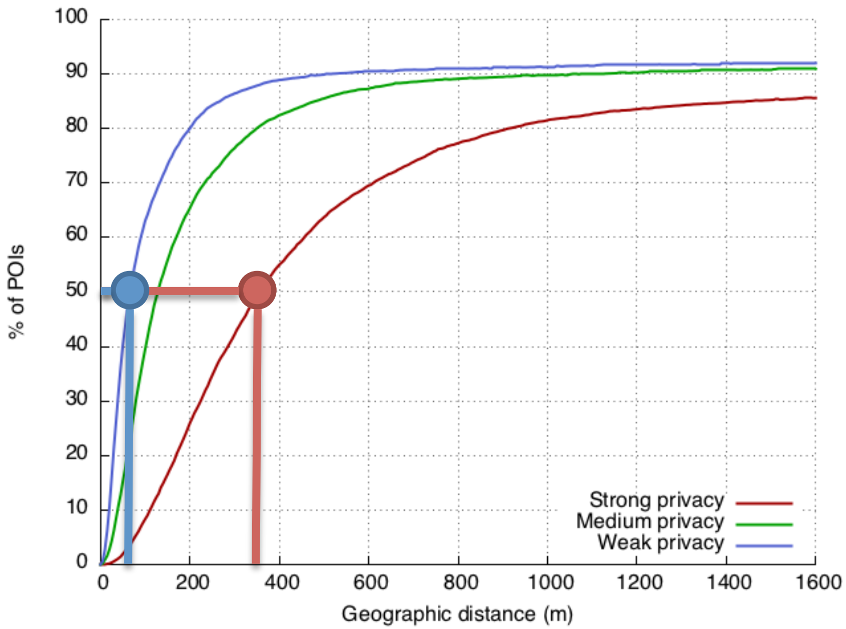
	SF cabs	Geolife
Reference (unobfuscated)	***1111 POIs*** (~ 2/user)	***258 POIs*** (~ 4/user)

Geographic distance

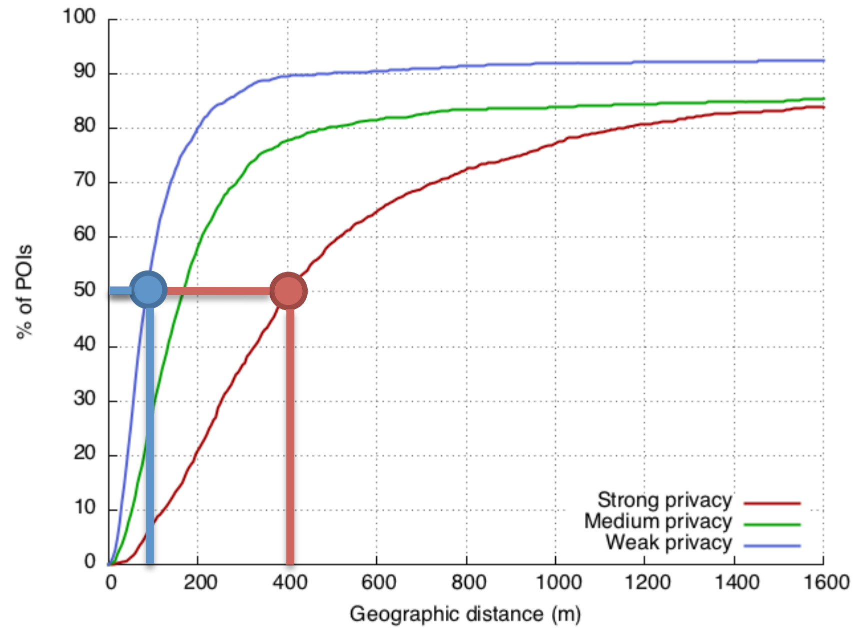
Geographic distance between an obfuscated POI and the nearest real POI



Cumulative geographic distance



SF cabs



Geolife

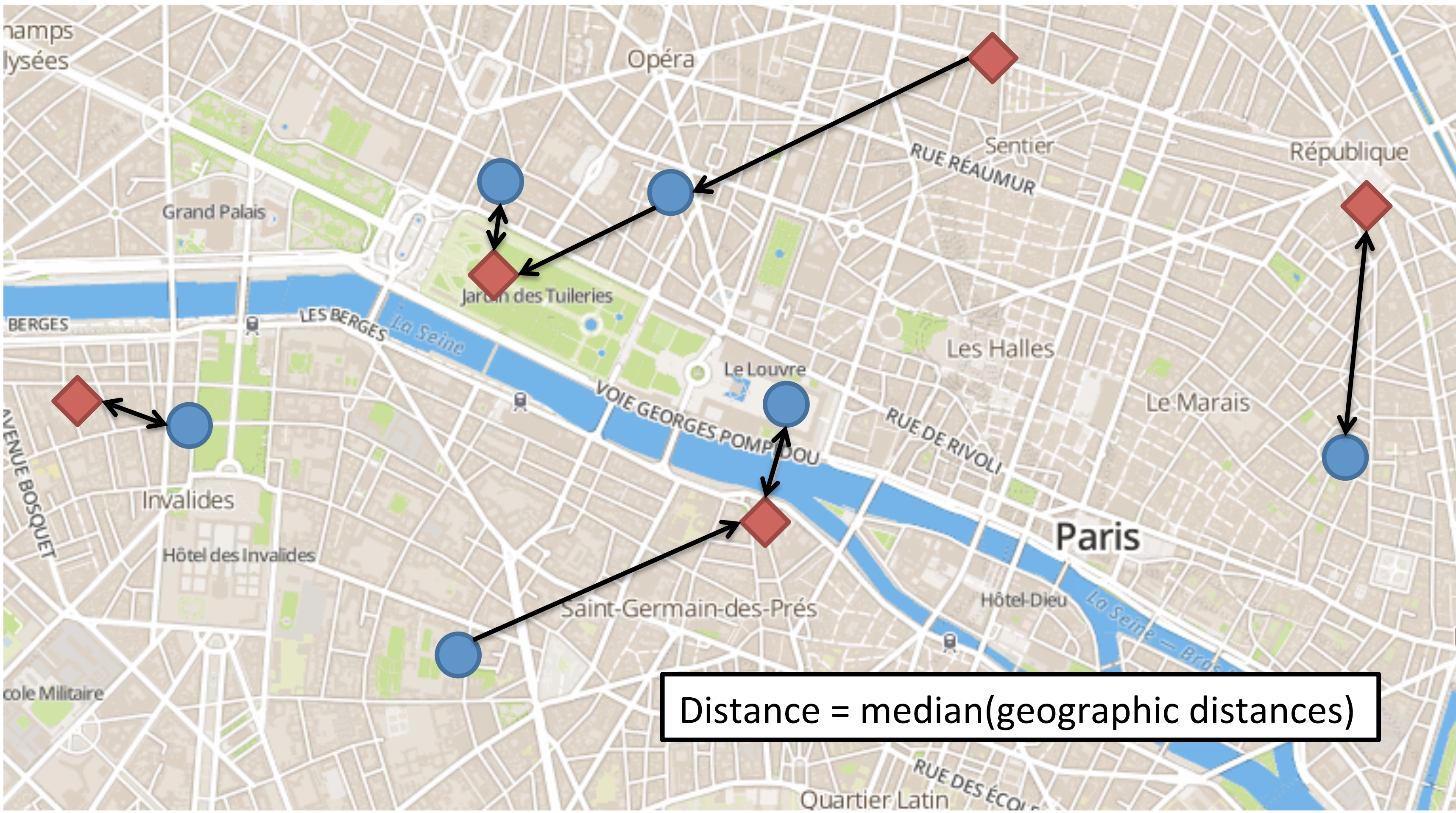
Re-identification rate

Scenario: I use a LBS without any protection and one day, I use a geo-indistinguishable mechanism.

Will my privacy be preserved or will the LBS be able to link my obfuscated trace with my original trace?

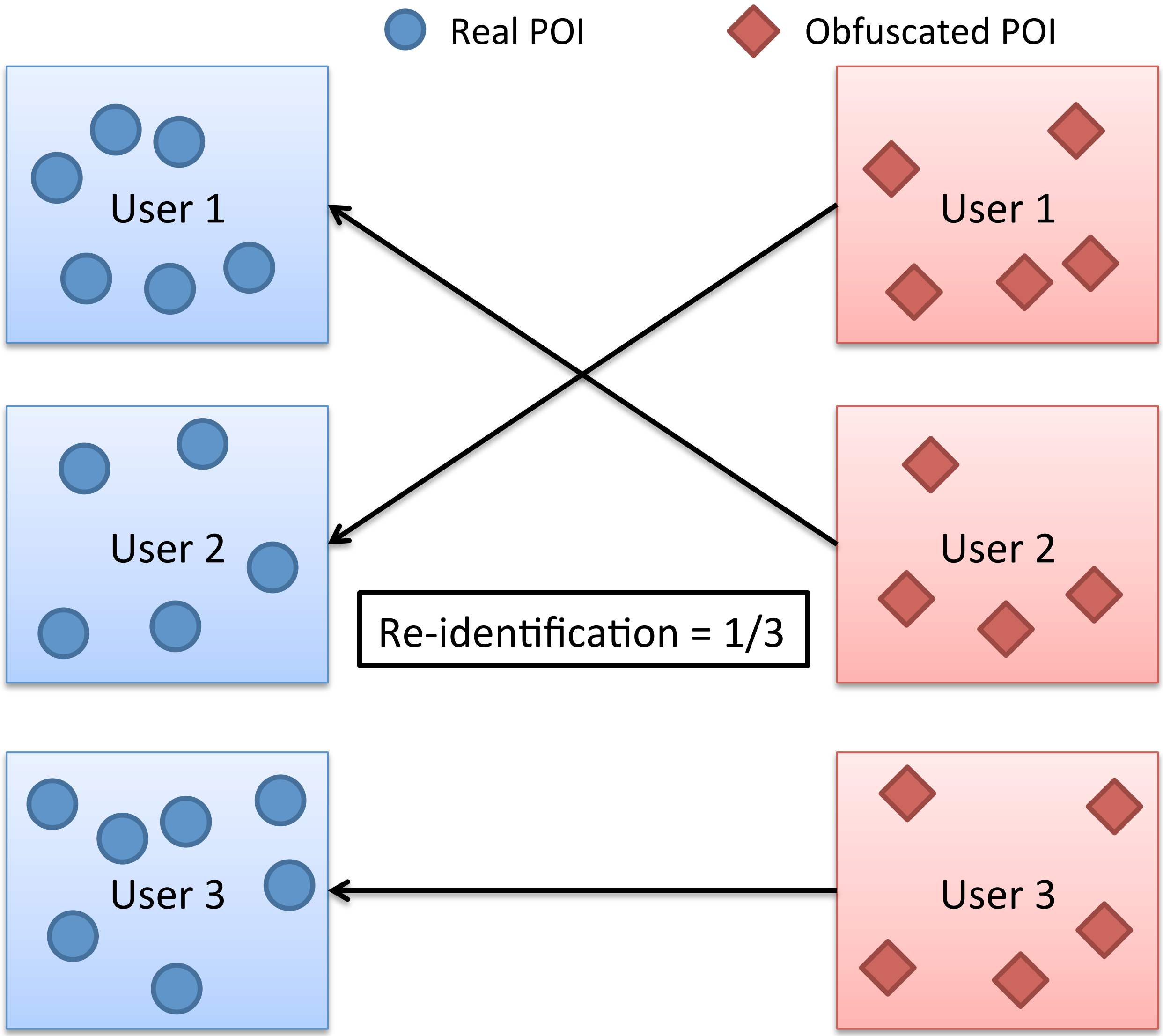
Re-identification rate

Real POI   Obfuscated POI



Re-identification rate

Associate to each set of obfuscated POIs the set of real POIs with which it has the minimal distance.



Re-identification rate

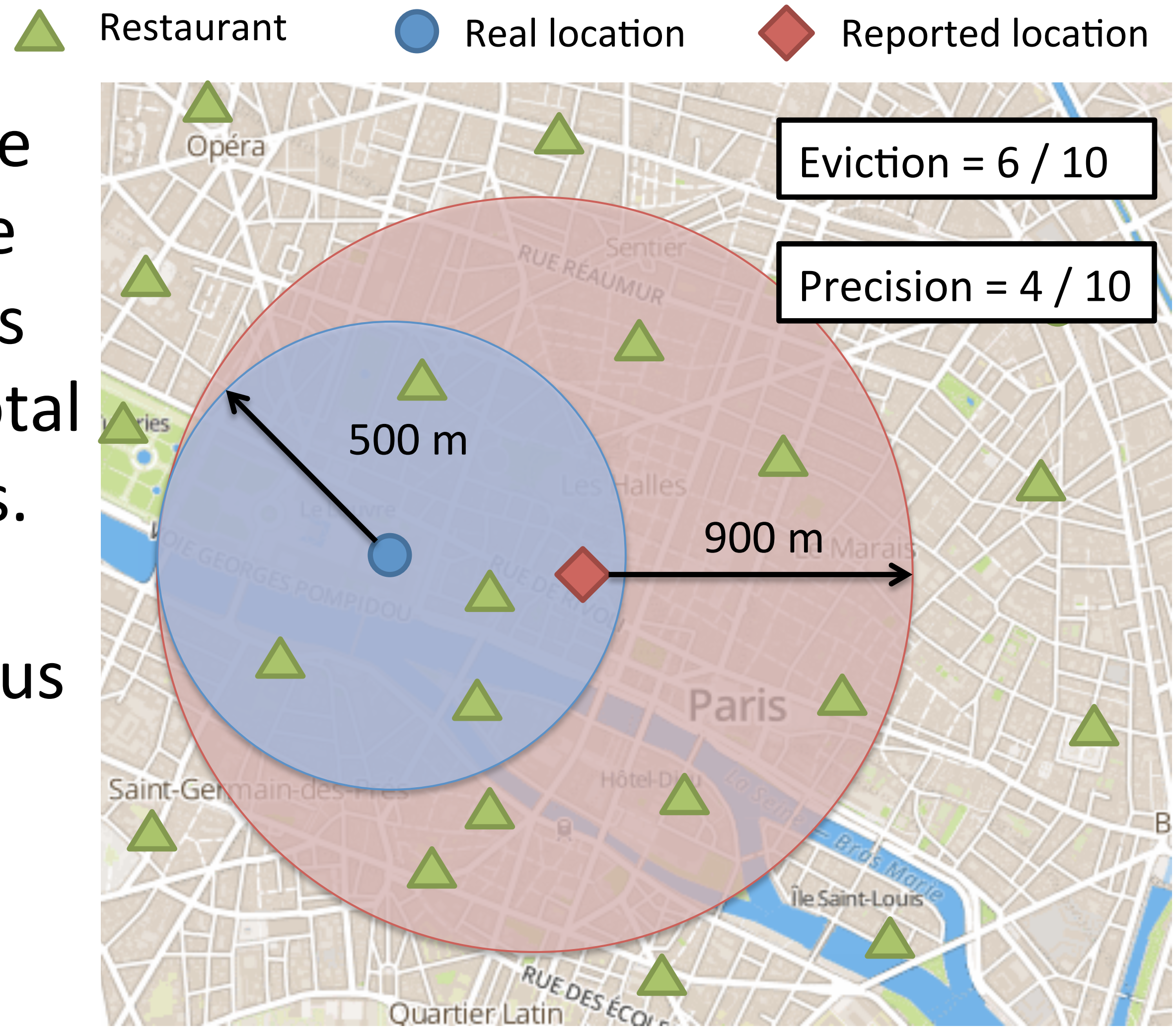
	SF cabs	Geolife
Strong privacy	6 %	63 %
Medium privacy	8 %	83 %
Weak privacy	10 %	90 %

- Few unique patterns in SF cabs data set, drivers are likely to have a similar behavior.
- Mobility patterns can be captured in Geolife and act like a fingerprint.

Measuring precision

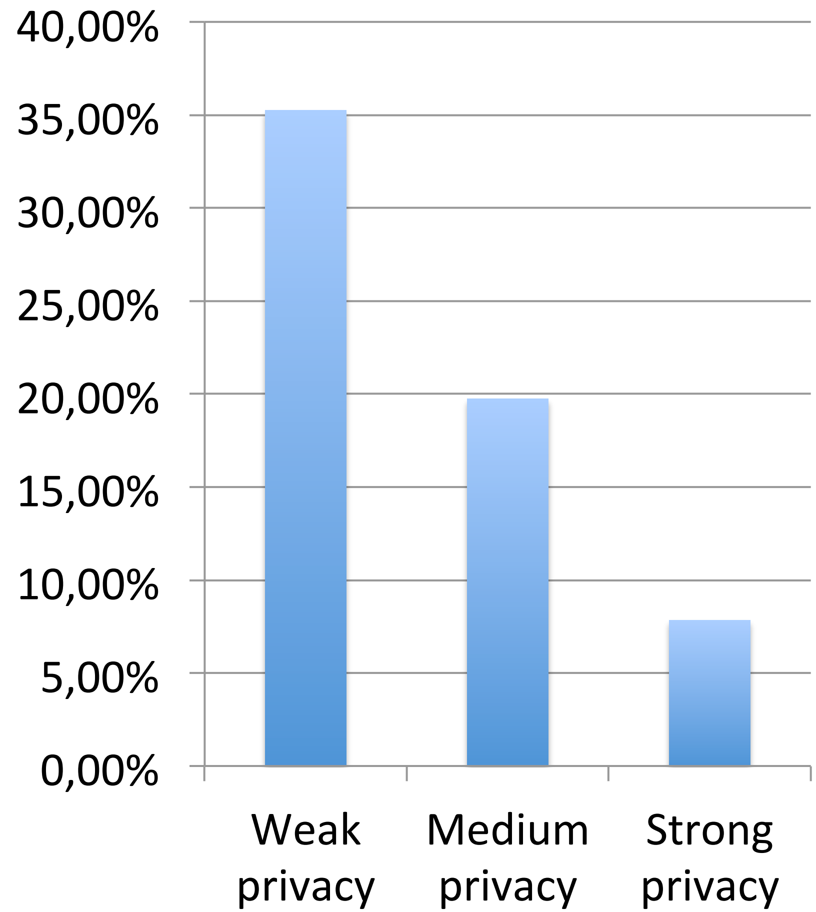
Eviction rate is the ratio between the number of useless results and the total number of results.

Precision is 1 minus the eviction rate.



Precision of results when querying LBS

- 100 points sampled from the SF cabs dataset
- Use a "*find restaurants 500 meters around me*" query against OpenStreetMap data



Outline

- Introduction
- About points of interest
- Protection mechanisms
- Experimental settings
- Evaluation metrics & results
- **Sum-up**

Conclusion

- Protection mechanisms improve privacy…
  - but still allow to infer a large quantity of sensitive information (> 60 %)
  - at the cost of degraded performance

- Difficult to achieve a trade-off between precision, utility and performance

Future work

- Study the exact impact of the temporal component
- Investigate if dynamically adapting the privacy parameter can help
- Propose counter-measures w.r.t. our framework and related work

Questions?