Time Distortion Anonymization for the Publication of Mobility Data with High Utility

Vincent Primault, Sonia Ben Mokhtar,
Cédric Lauradoux and Lionel Brunie

Mobility data usefulness…

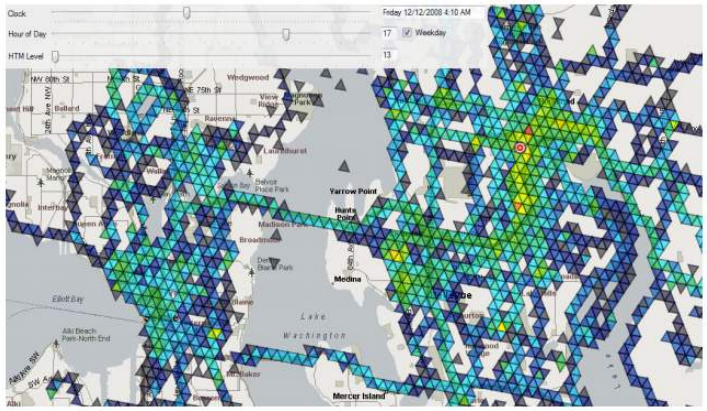
waze

apisense BETA

orange™

Companies collecting data

Real-time traffic, traffic prediction



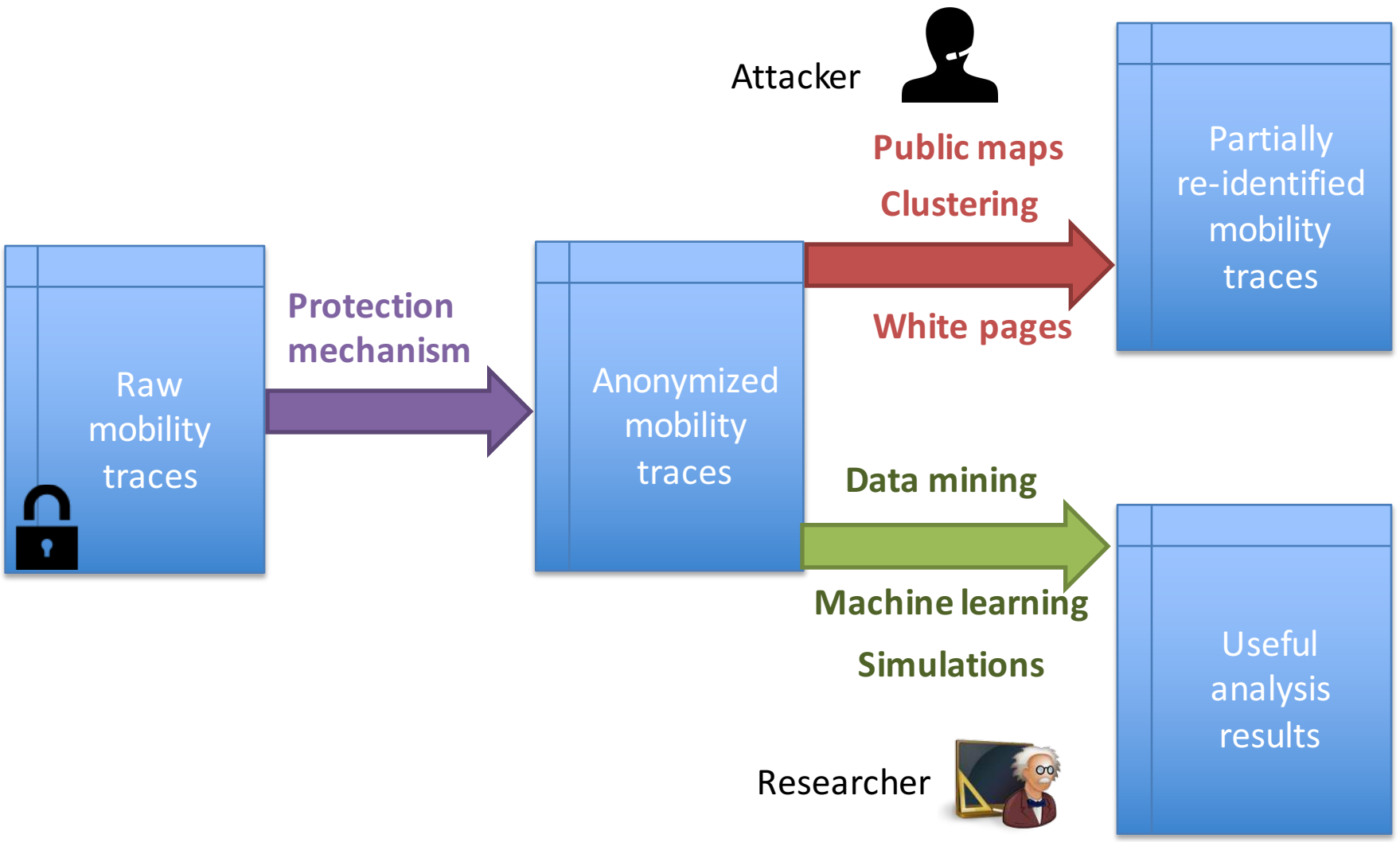
Long-term place prediction

… and threats

Gambs et al. **Show Me How You Move and I Will Tell You Who You Are.** *Transactions on Data Privacy, 2011.*

Google | google.com/datacenters

Privacy-preserving data publication



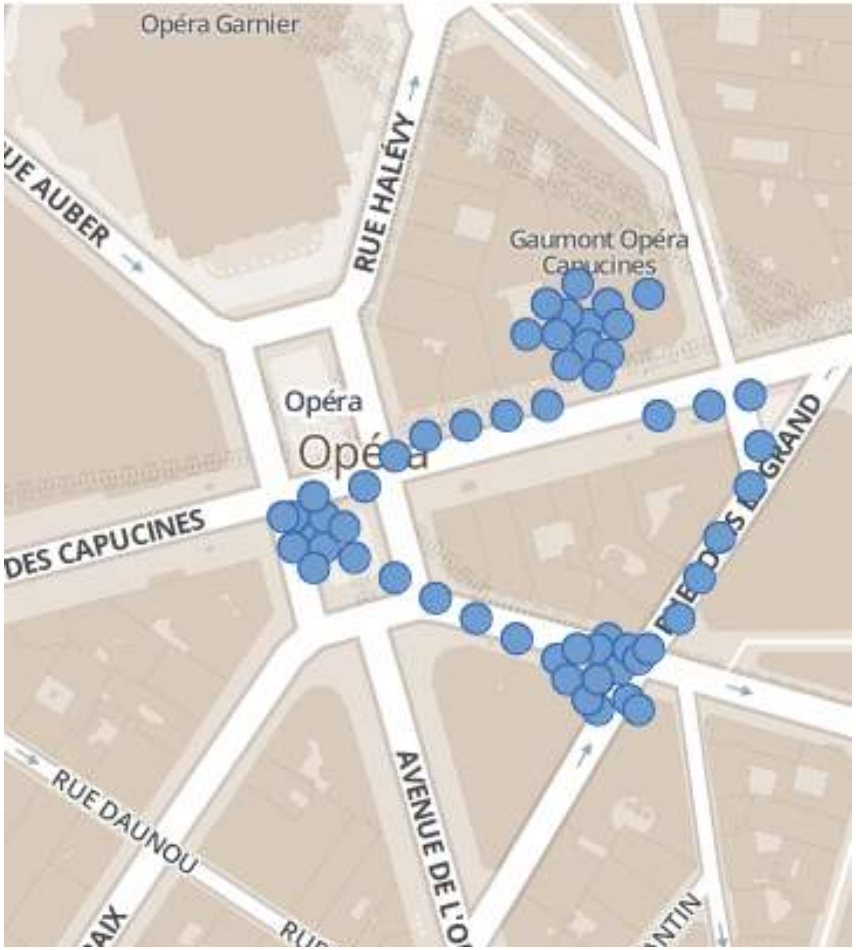
Outline

- Introduction
- **State of the art**
- Making a PROMESSE
- Experimental evaluation
- Conclusion

A mobility trace

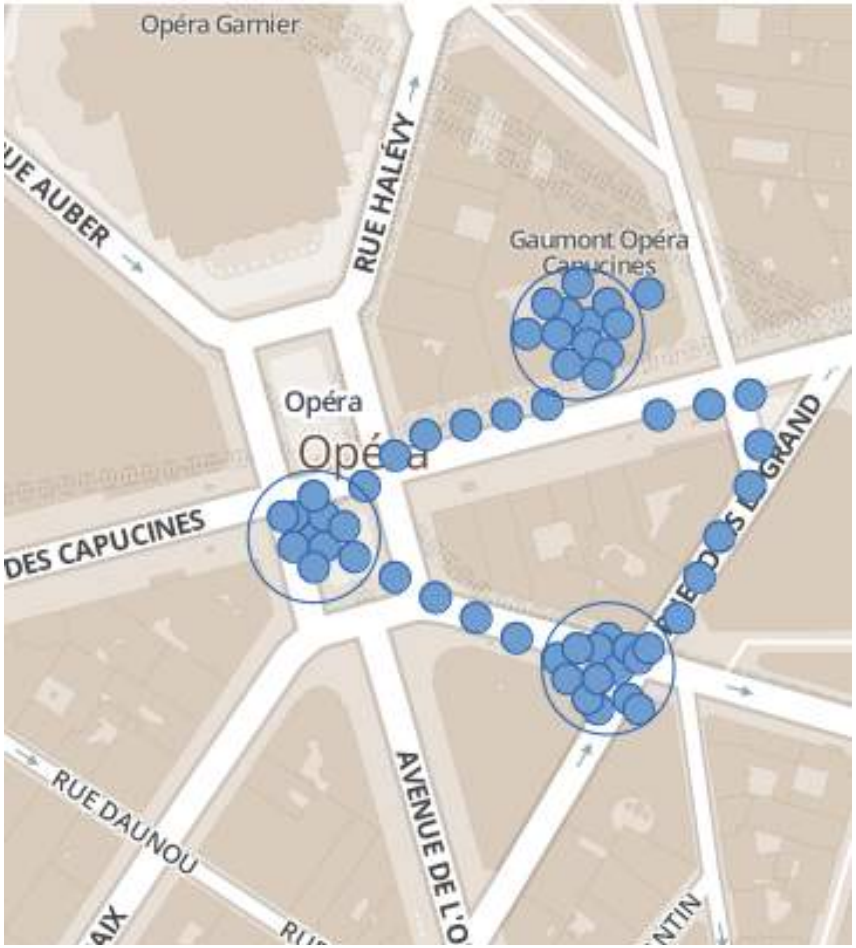
A trace is a temporally ordered list of records belonging to a same user.

A record is a triplet (user, location, timestamp).



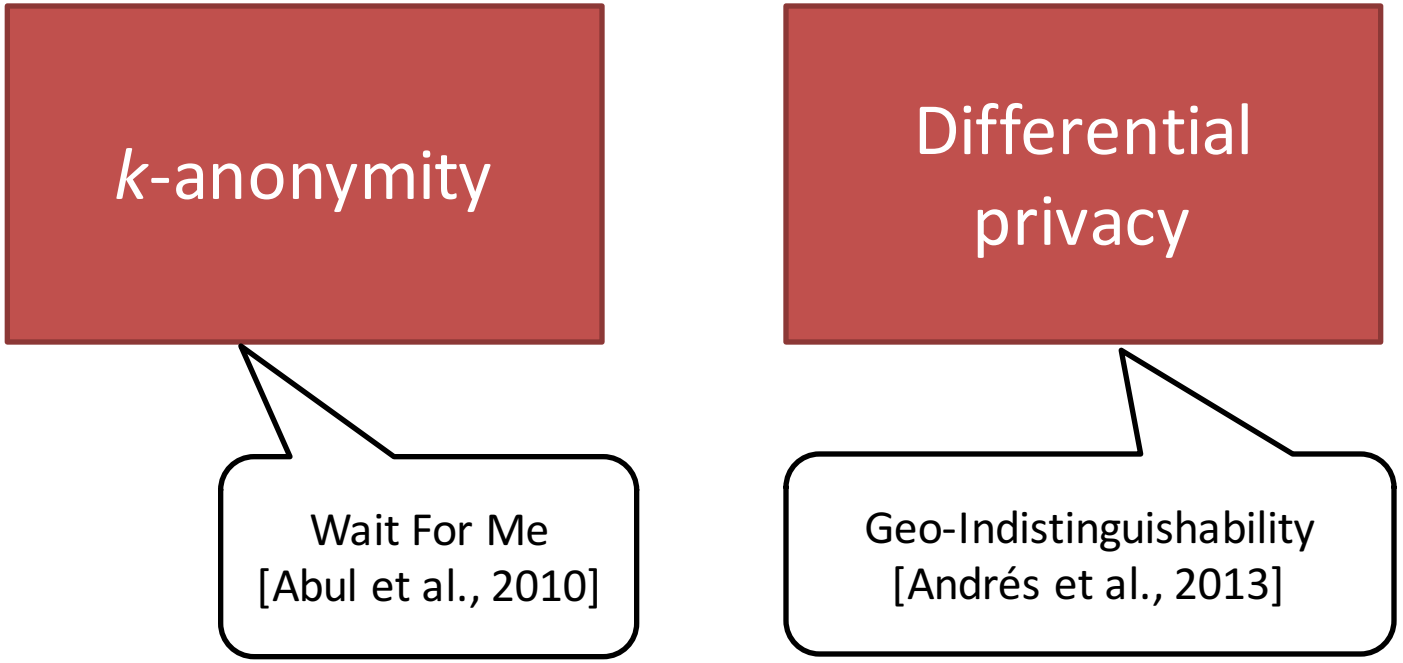
Extraction of points of interest (POIs)

Done by using, e.g., an appropriate clustering algorithm.



Points of interest convey semantic information about habits and can lead to users re-identification.

Location privacy protection mechanisms for data publication

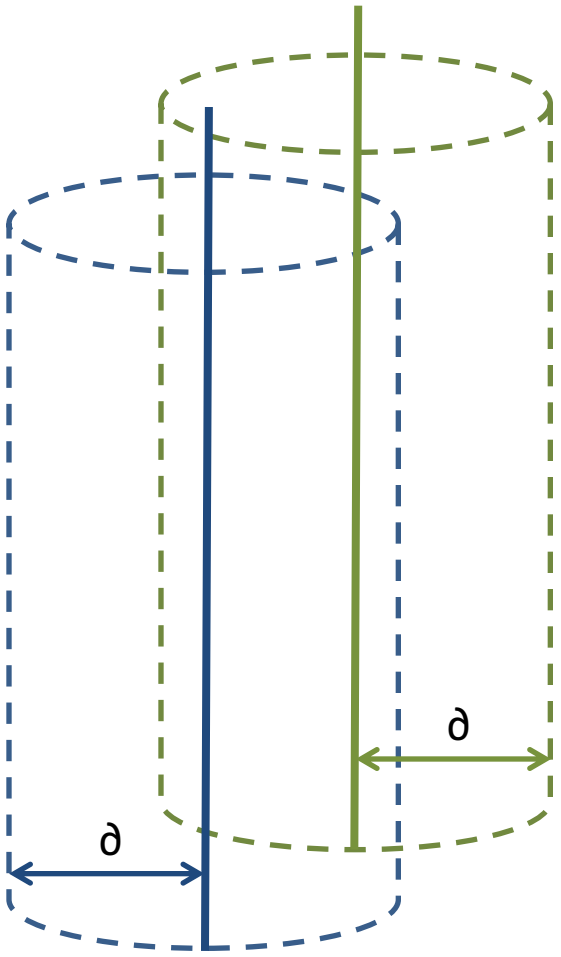


Abul et al. **Anonymization of moving objects databases by clustering and perturbation**. *Information Systems, 2010.*
Andrés et al. **Geo-indistinguishability: Differential privacy for Location-based Systems**. *CCS, 2013.*

Wait For Me

ð represents the incertitude that comes from GPS measurements.

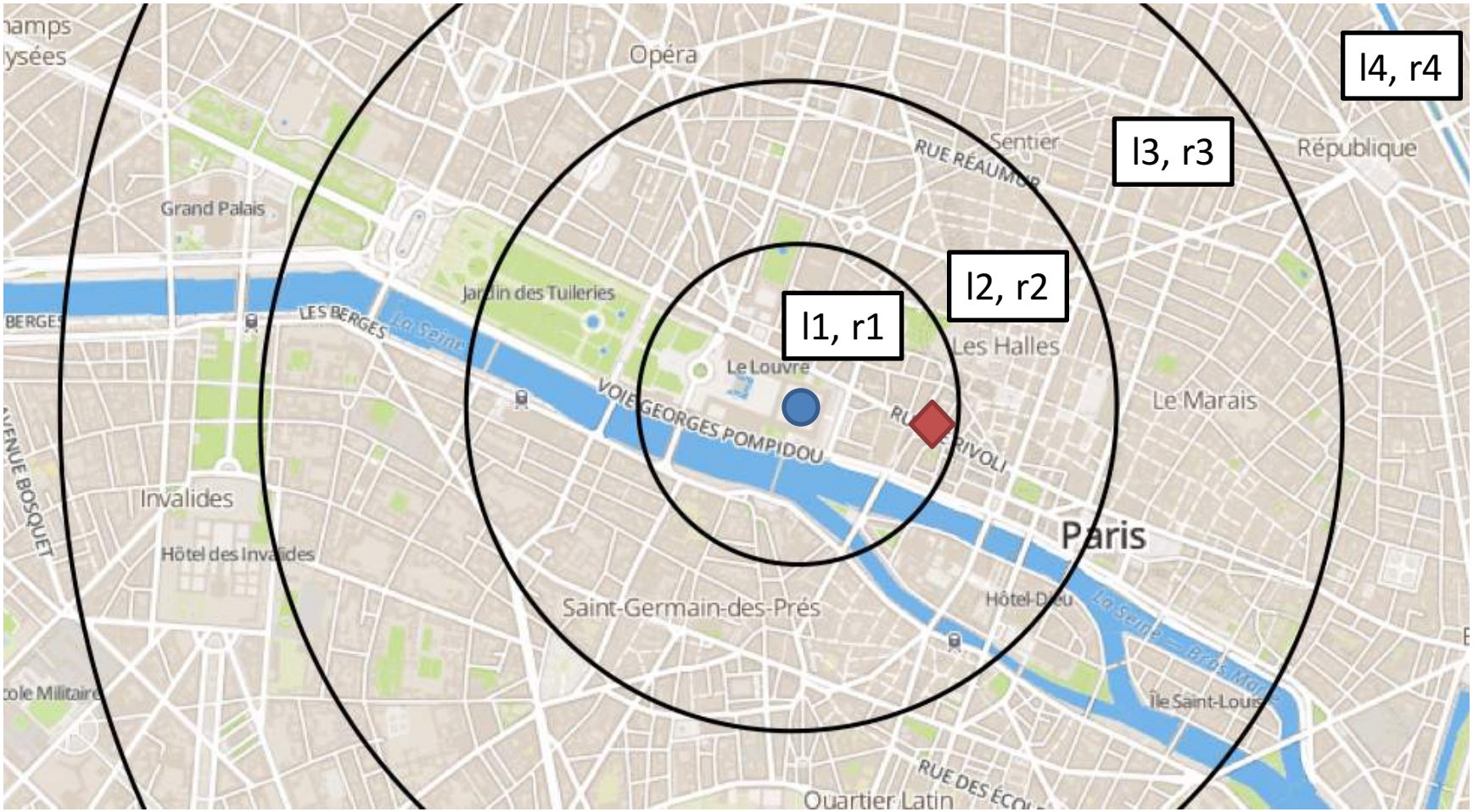
Wait For Me enforces (*k*,ð)-anonymity, i.e., there is always at least *k* users in a cylinder of radius ð/2.



Abul et al. **Anonymization of moving objects databases by clustering and perturbation**. *Information Systems, 2010.*

Geo-Indistinguishability

Level of privacy l_i within r_i proportional to an ε Real location Protected location



Andrés et al. **Geo-indistinguishability: Differential privacy for Location-based Systems**. *CCS, 2013.*

Outline

- Introduction
- State of the art
- **Making a PROMESSE**
- Experimental evaluation
- Conclusion

Intuition behind our work

No state-of-the-art mechanism is both privacy-preserving and useful for data scientists.

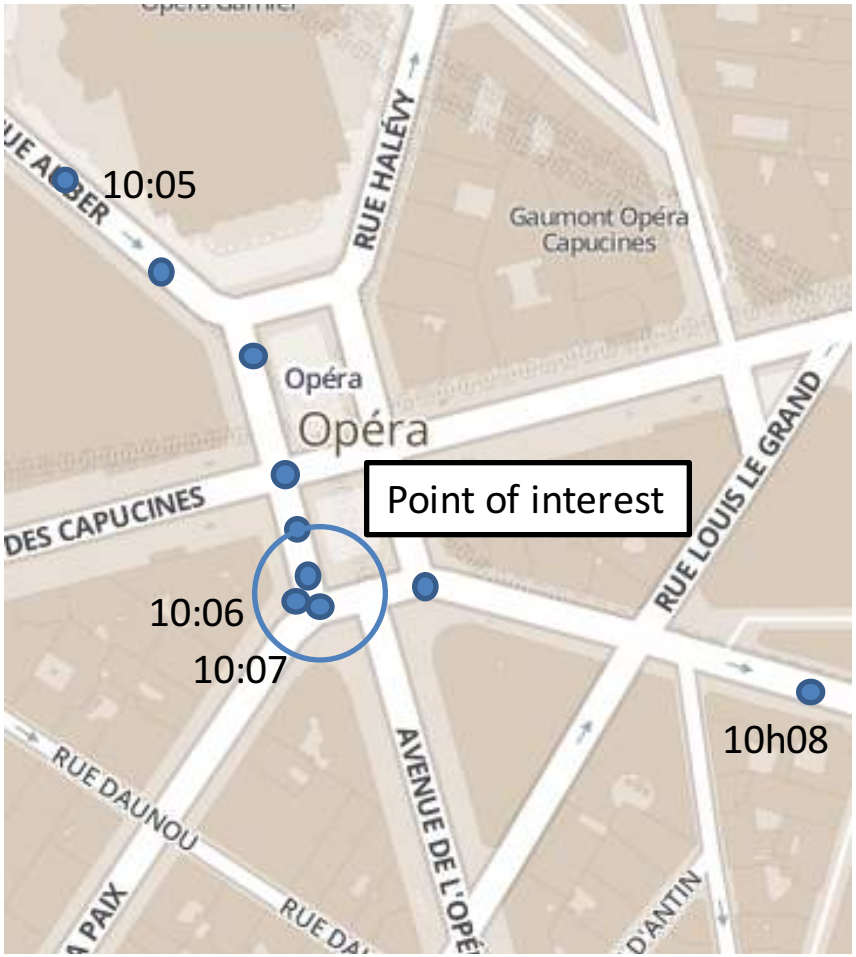
Almost all of them alter the geographical information in some way.

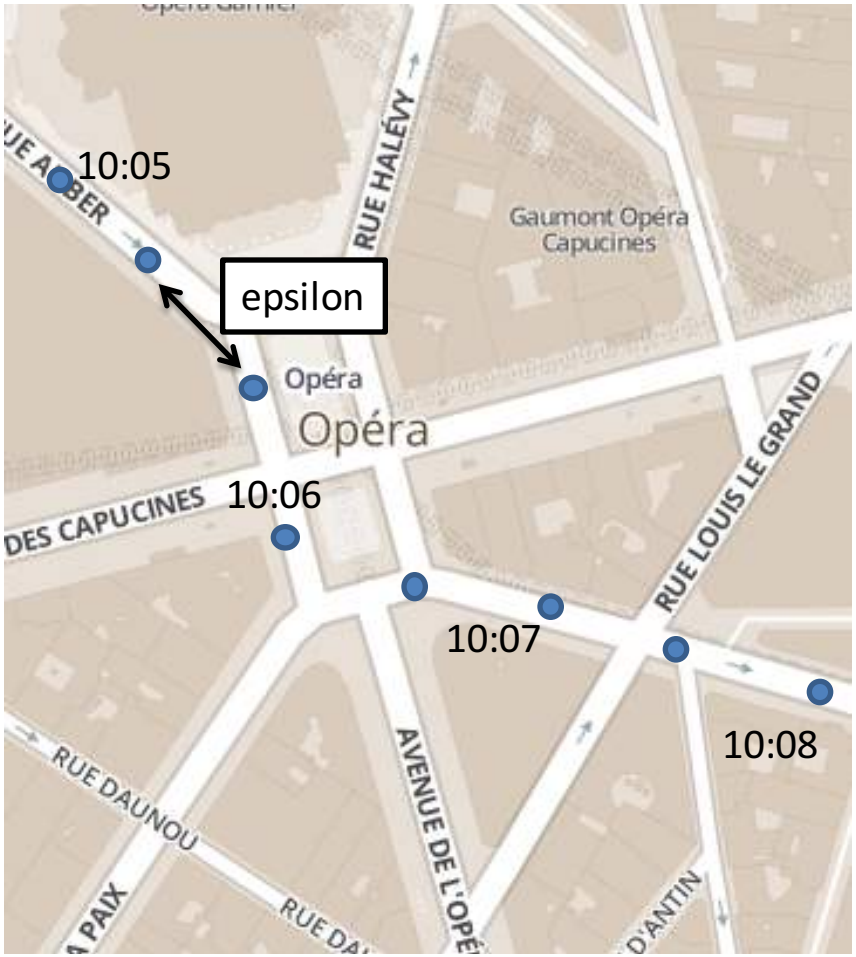
We believe geographical information is the most important one, so we propose a new mechanism that minimally distort the location.

Hiding POIs with speed smoothing

The idea	How?
To guarantee a constant speed along a trace. More challenging to identify where a user stops, and therefore her POIs.	Divide traces into smaller trajectories, typically one day long. Enforce an equal duration and length between two consecutive records.

Speed smoothing





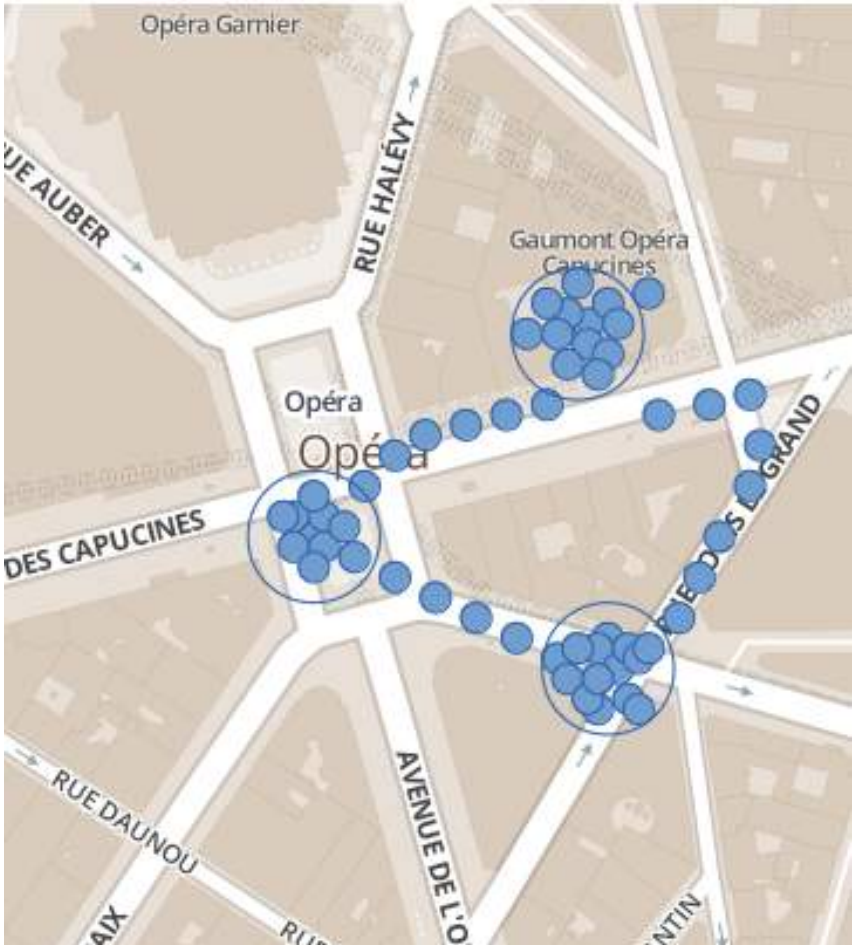
Outline

- Introduction
- State of the art
- Making a PROMESSE
- **Experimental evaluation**
- Conclusion

Experimenting with three real-life datasets

	Cabspotting	Geolife	MDC
Records	8,9M	3,8M	1,1M
Traces	5,5k	2,4k	4,6k
Avg trace duration	32 h	3 h	3 h
Avg sampling rate	72 s	7 s	32 s

POIs retrieval

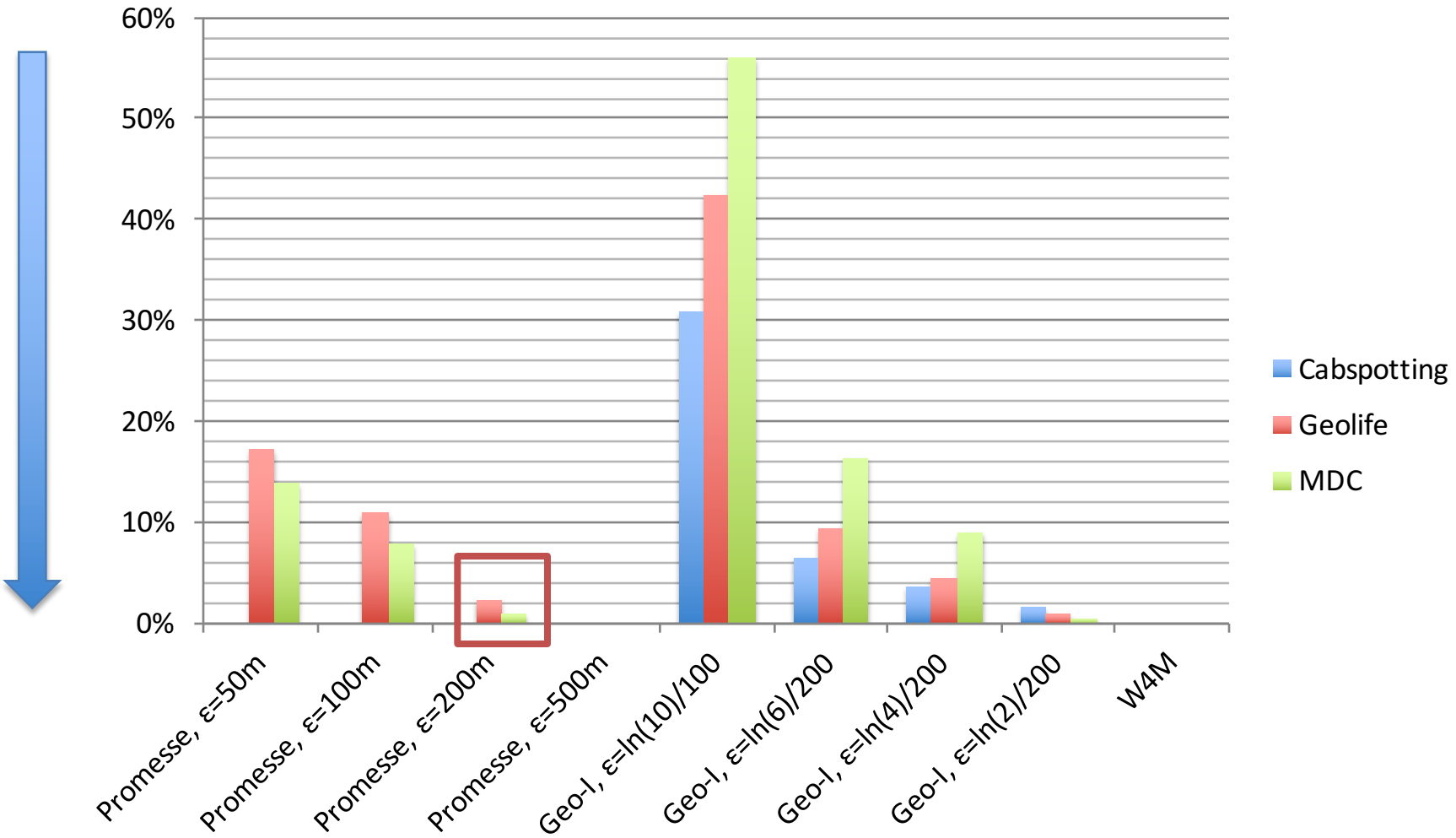


POIs with maximum diameter of **200 meters** and minimum duration of 15 minutes.

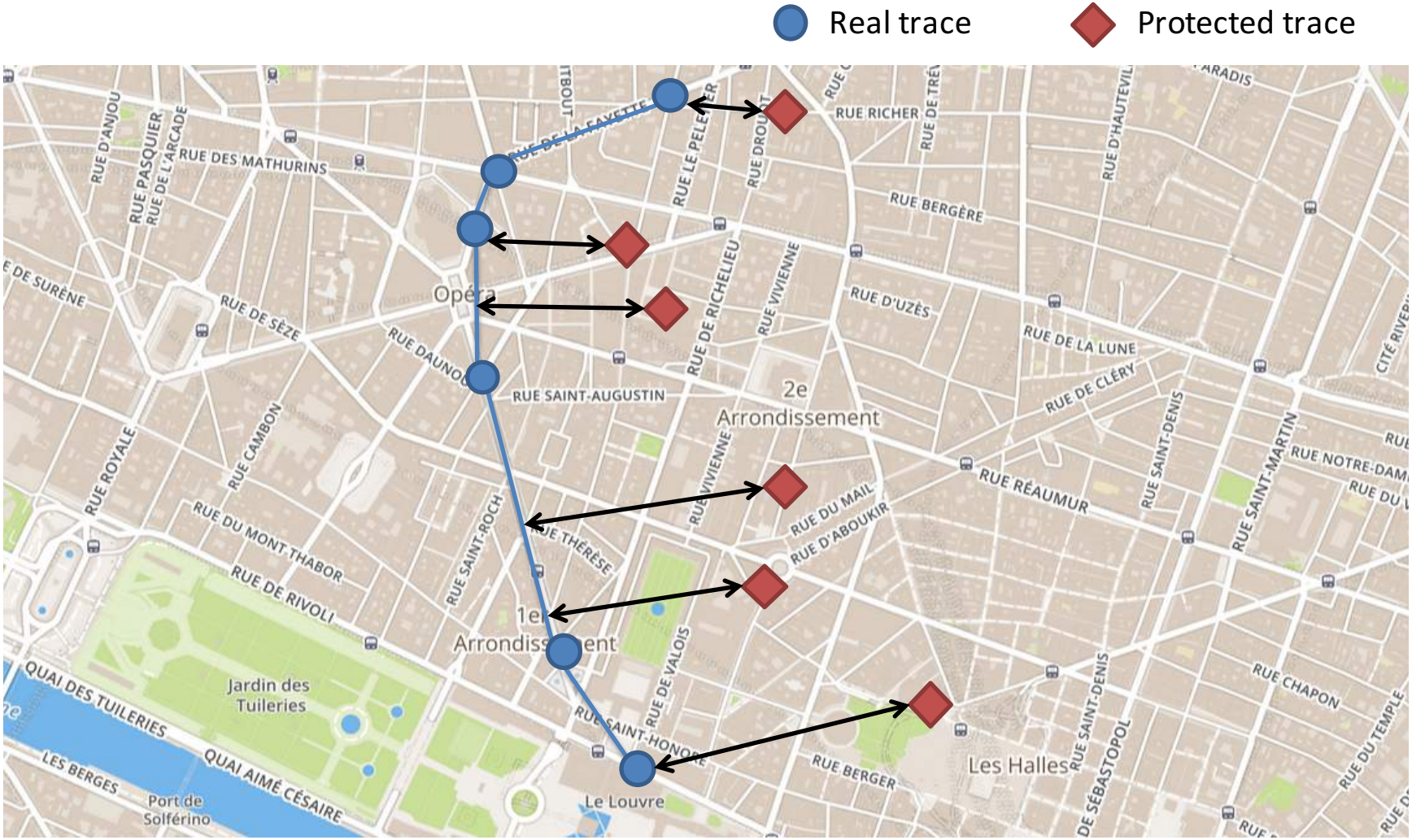
Two POIs match if their centroids are within 100 meters.

POIs retrieval (F-score)

Lower is better

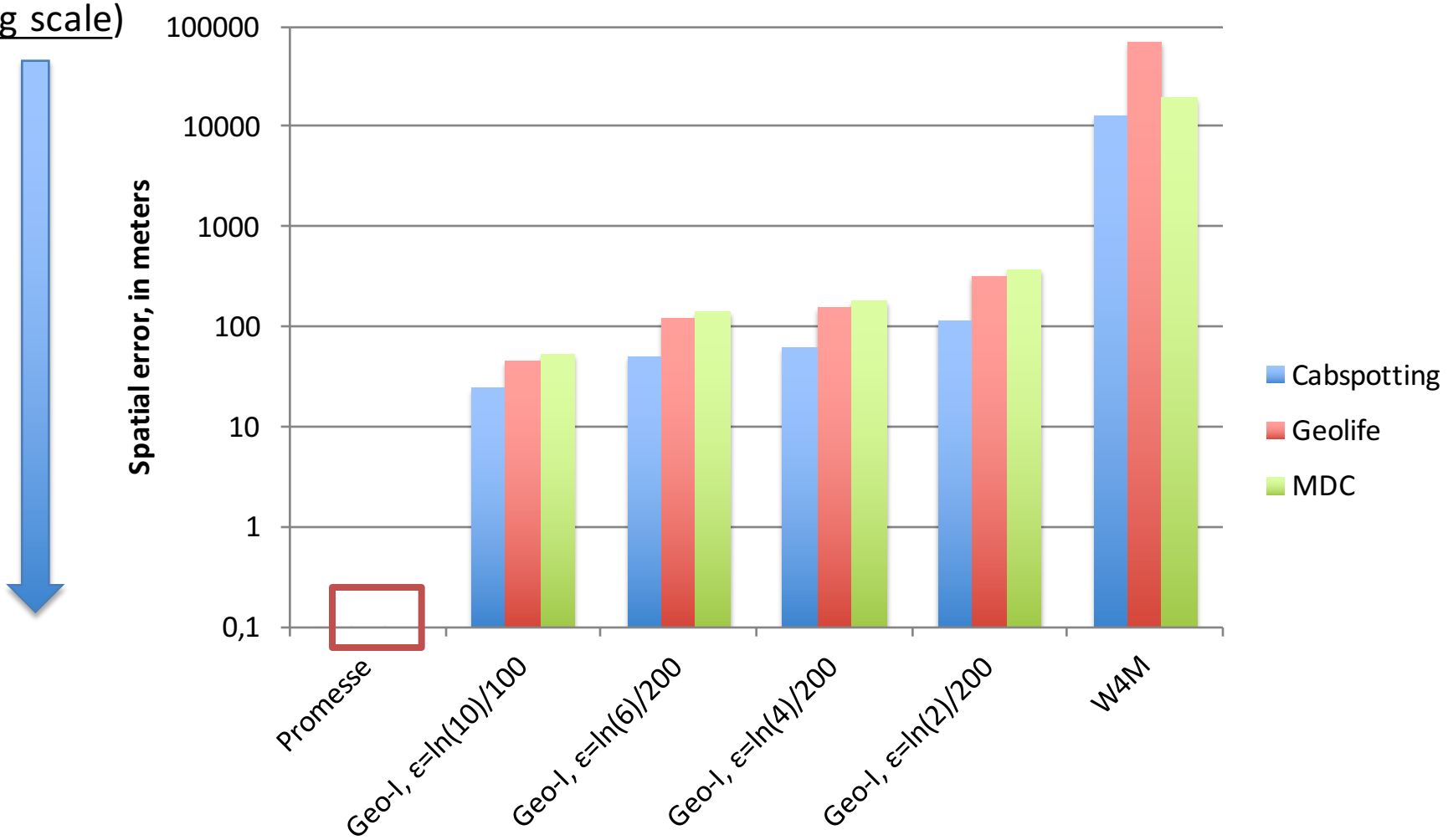


Average spatial error

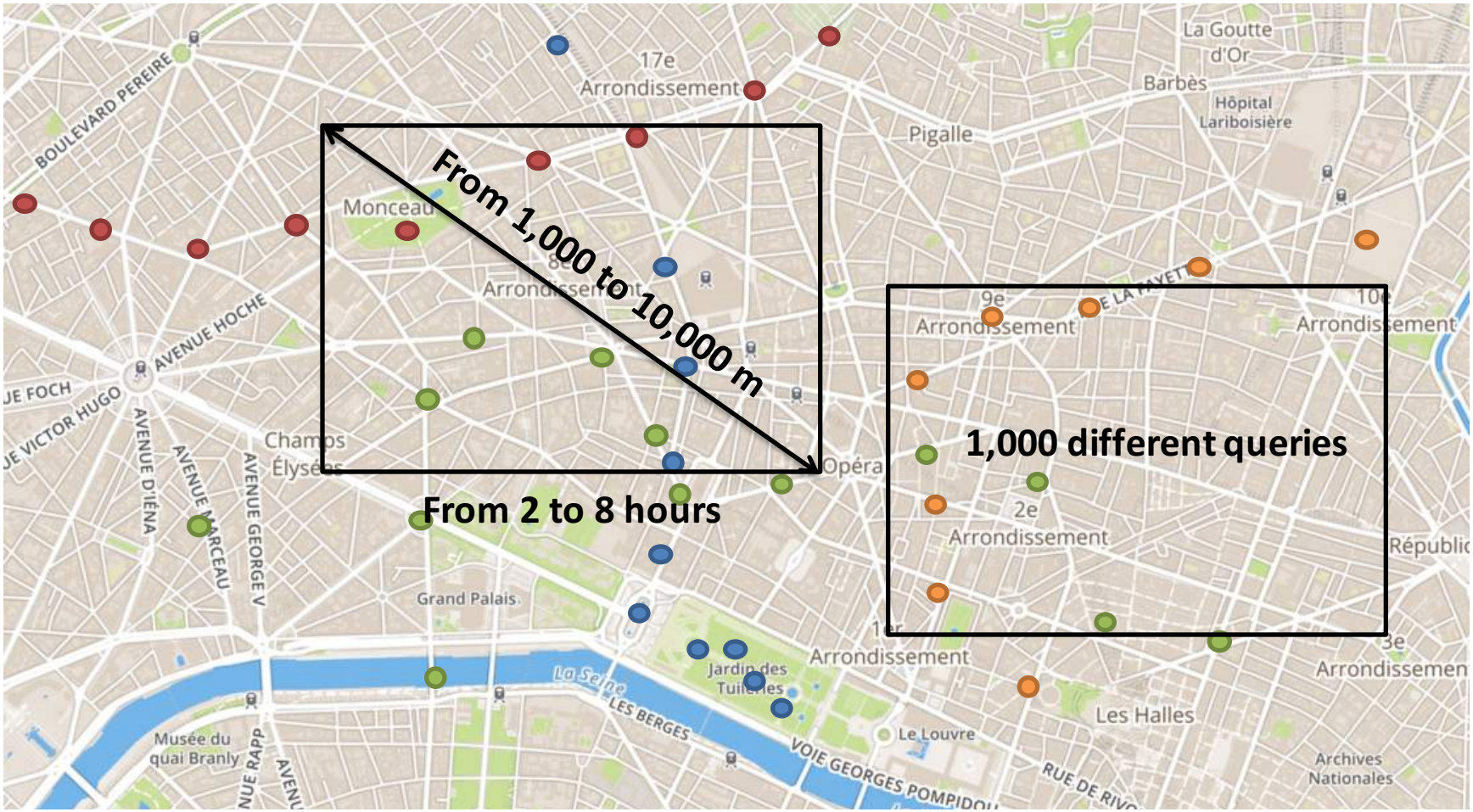


Average spatial error

Lower is better (log scale)



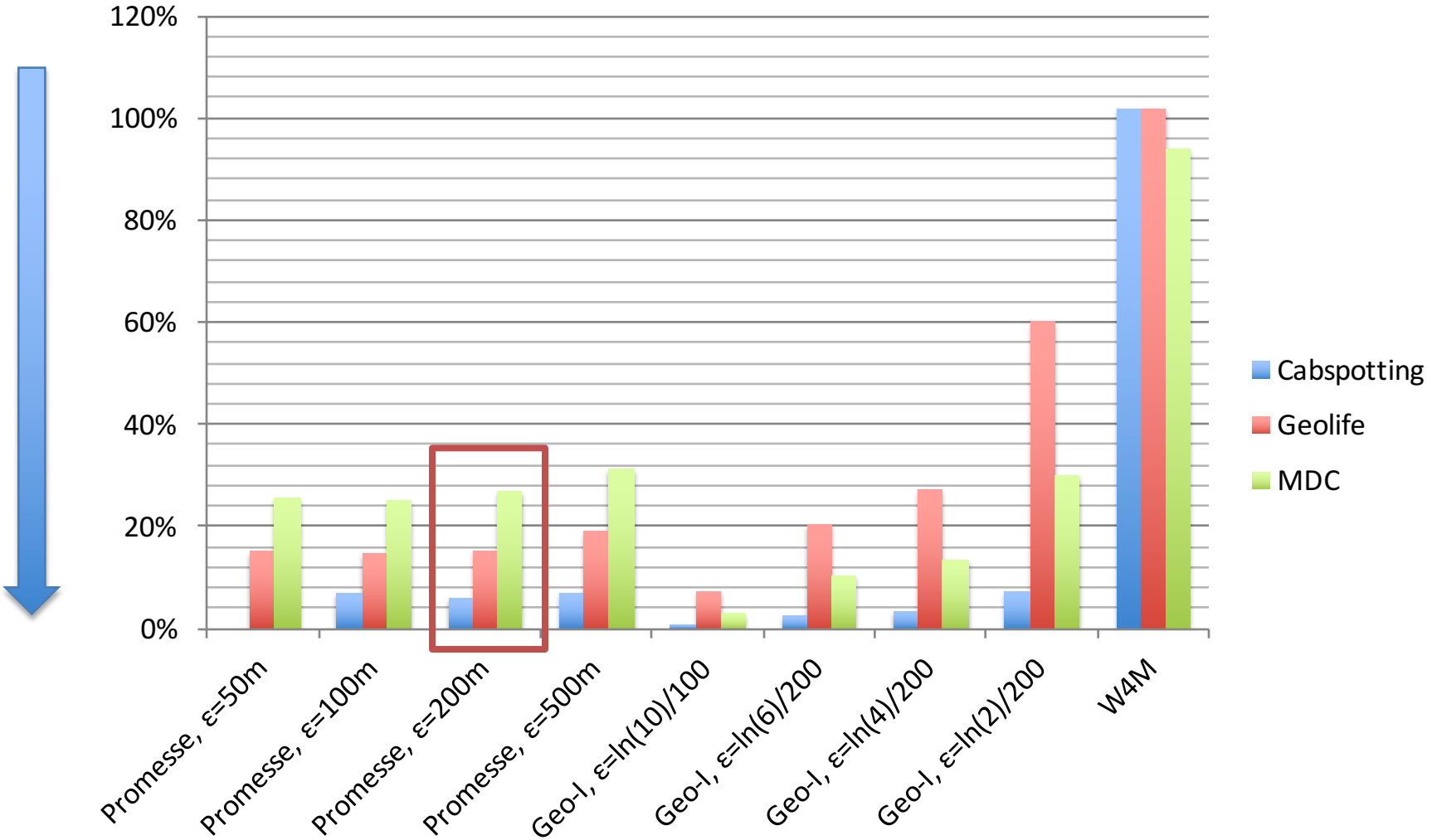
Range queries distortion



Distortion is $|Q(D) - Q(D')|/Q(D)$

Range queries distortion

Lower is better



Outline

- Introduction
- State of the art
- Making a PROMESSE
- Experimental evaluation
- Future work
- **Conclusion**

Summary

- Introduced time distortion, opened a new research direction.
- Implemented a new protection mechanism for data publishing, addressing a severe threat while maintaining high utility.
- Evaluated against three real-life datasets.

Questions