Geo-indistinguishability: A Principled Approach to Location Privacy

Kostas Chatzikokolakis
CNRS, INRIA, LIX Ecole Polytechnique

joint work with
Miguel Andrés, Nicolás Bordenabe,
Catuscia Palamidessi, Marco Stronati

PRINCESS QIF Day, Dec 16, 2014

Location-Based Systems

A **location-based system** is a system that uses geographical information in order to provide a service.

- Retrieval of Points of Interest (POIs).
- Mapping Applications.
- Deals and discounts applications.
- Location-Aware Social Networks.

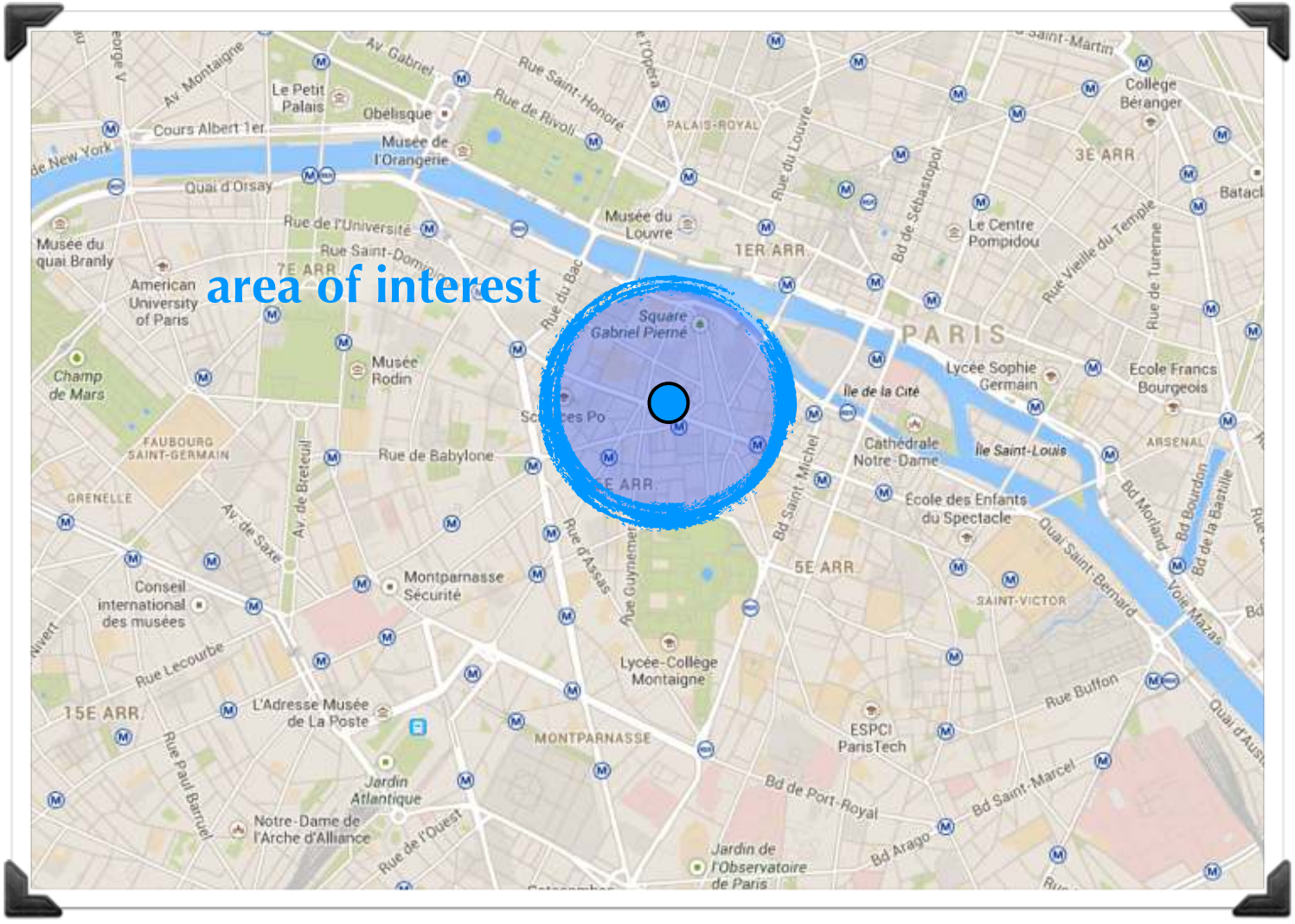
Location-Based Systems

- Location information is sensitive. (it can be linked to home, work, religion, political views, etc).
- Ideally: we want to **hide our true location**.
- Reality: we need to **disclose some information**.

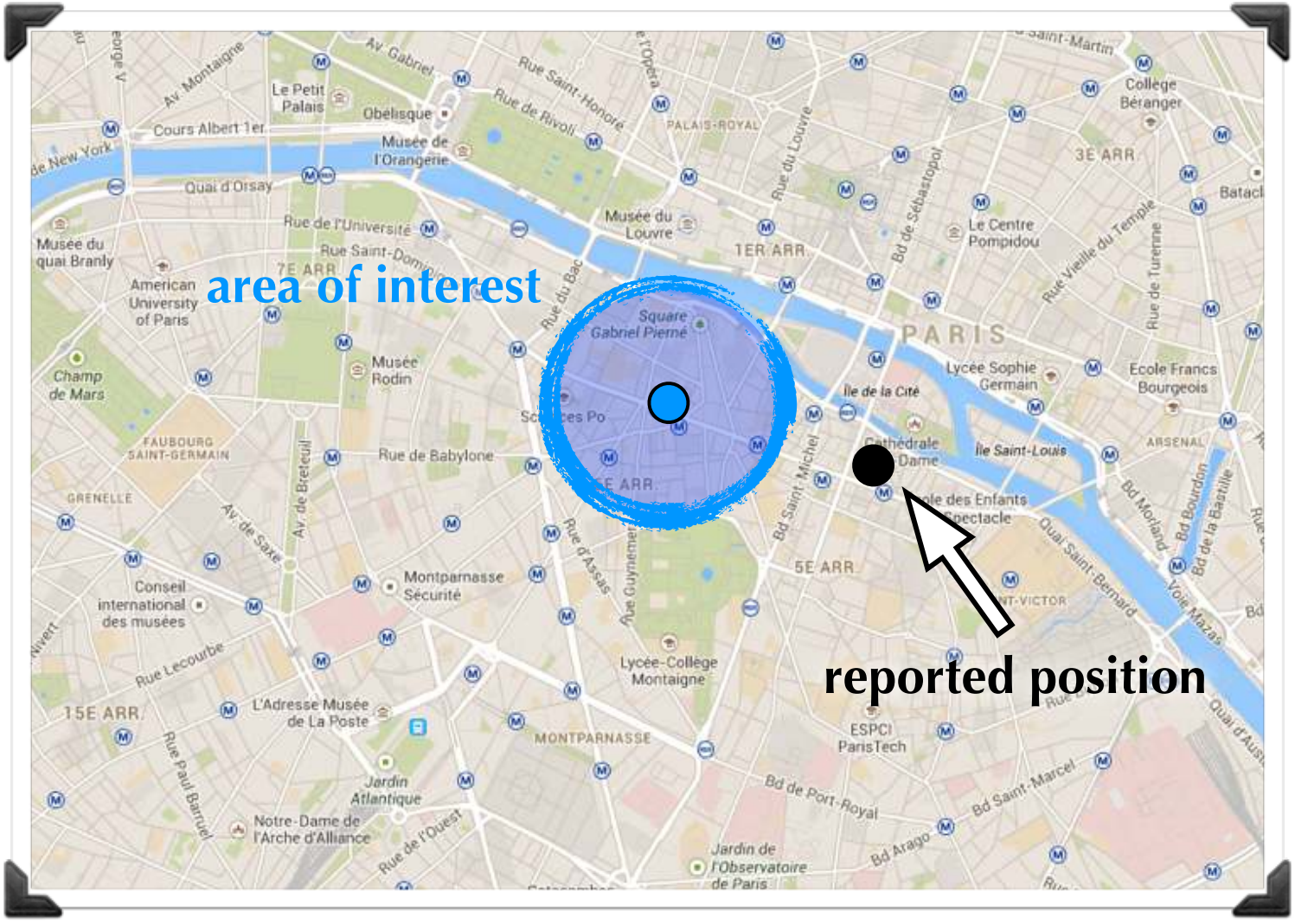
Example

- Find restaurants within 300 meters.
- Hide location, **not identity**.
- Provide **approximate location**.

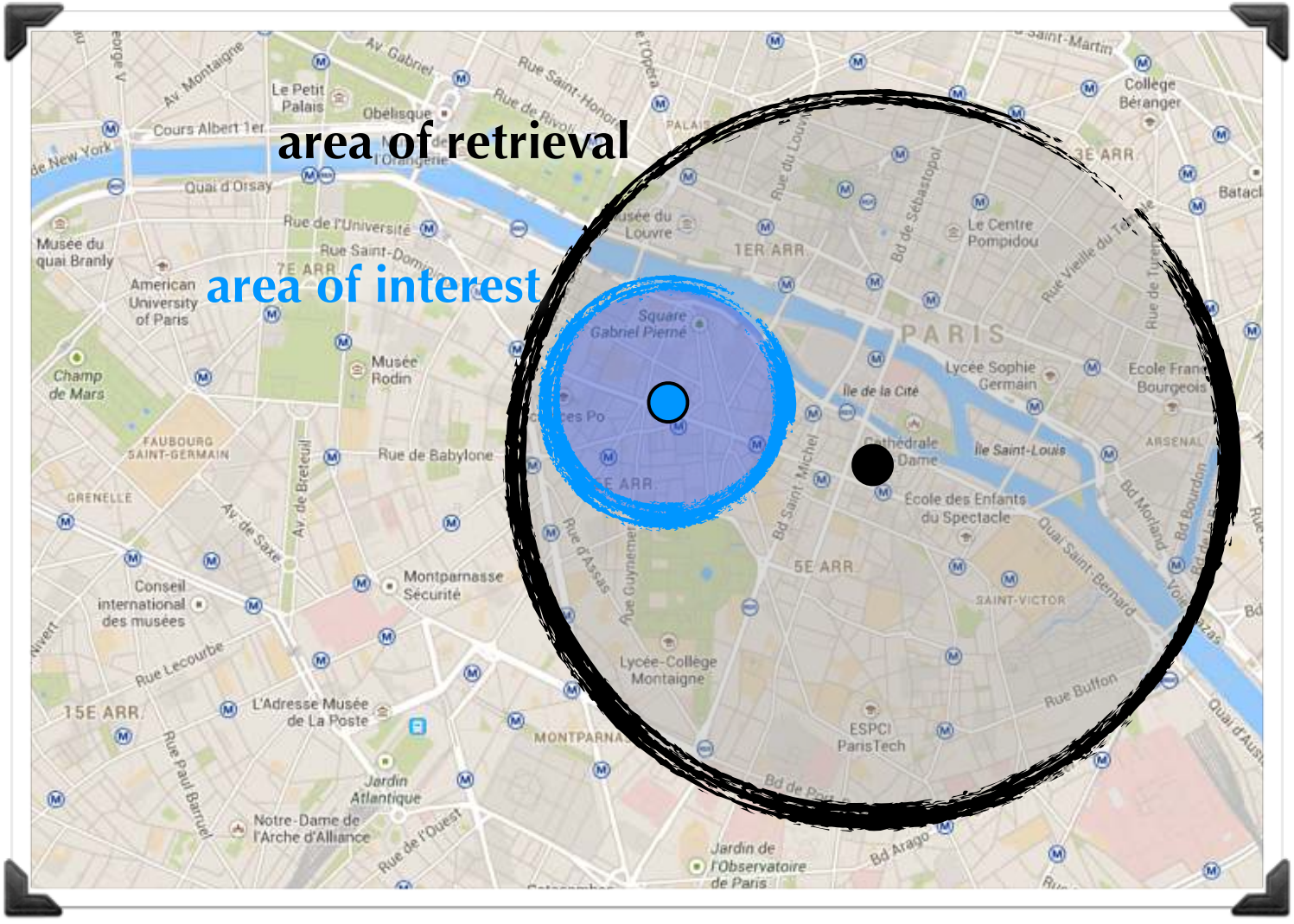
Obfuscation



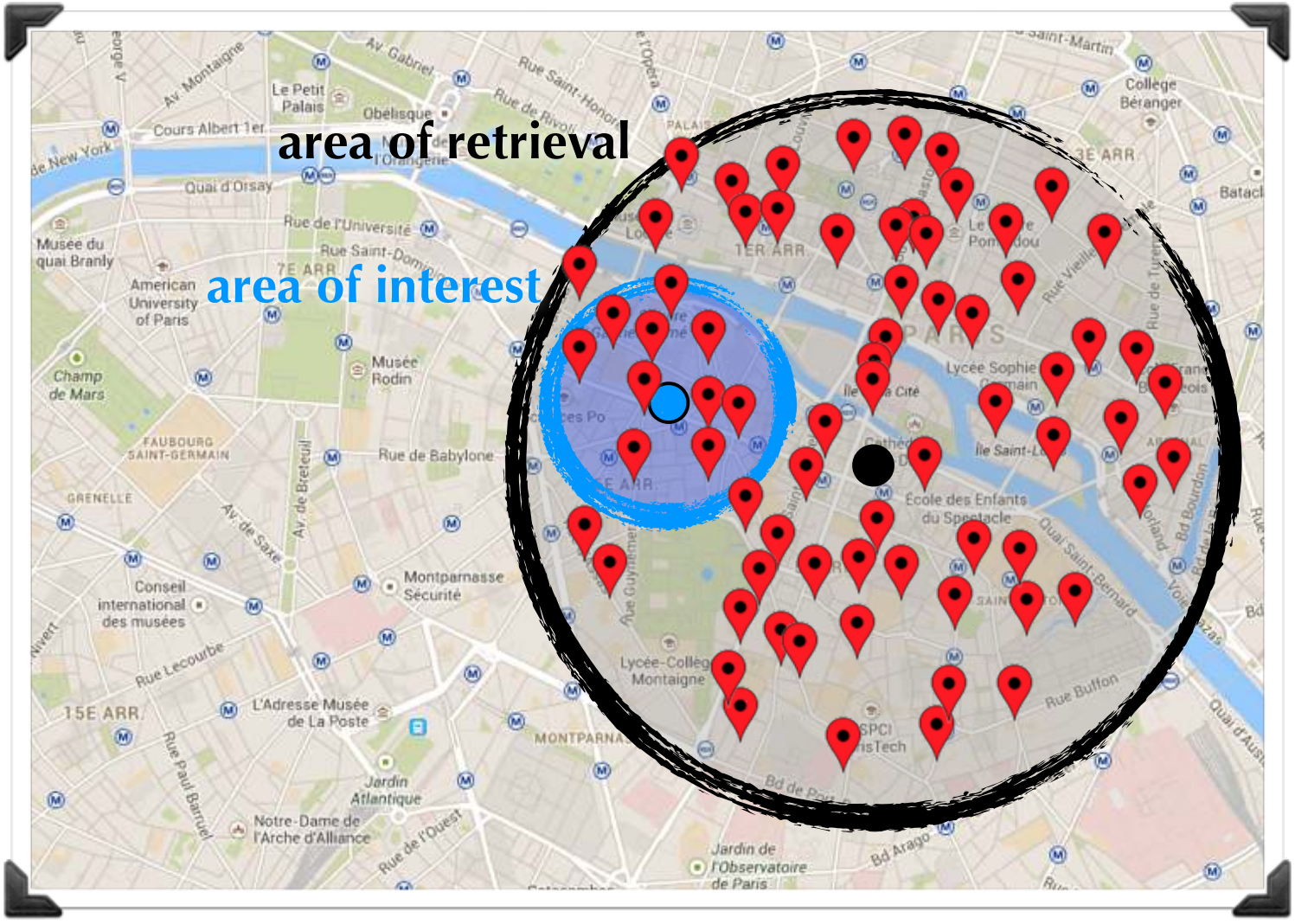
Obfuscation



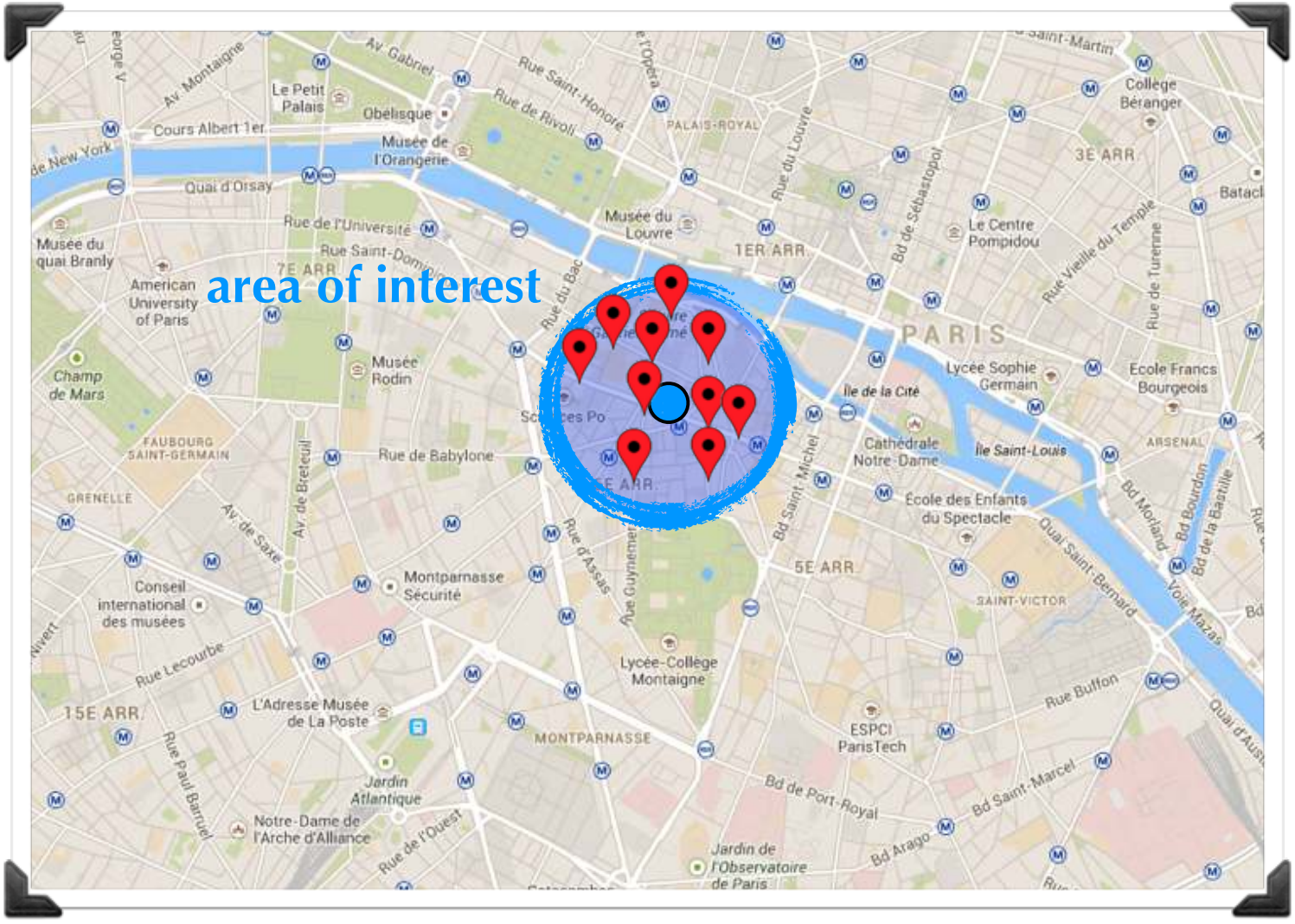
Obfuscation



Obfuscation



Obfuscation

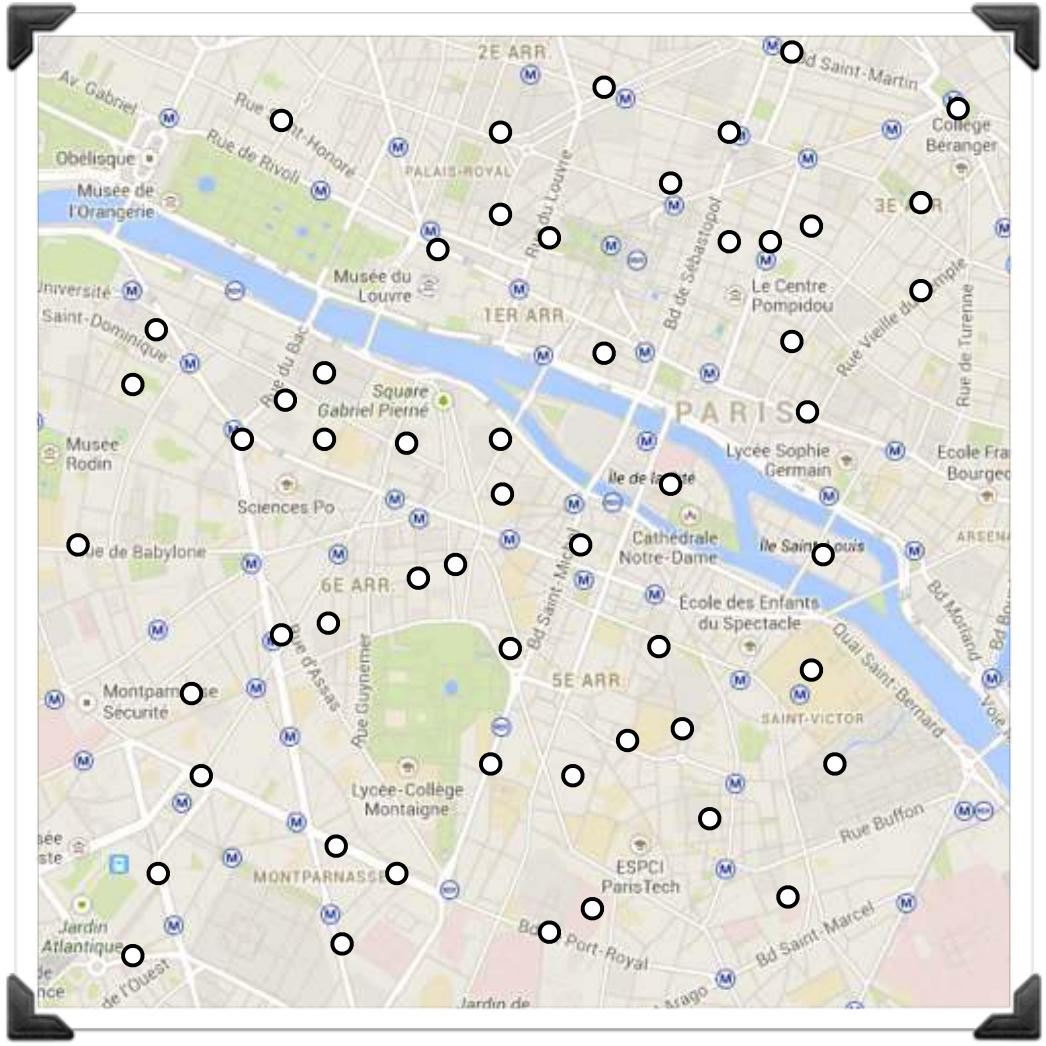


The Goals

- We want an **obfuscation mechanism**.
- Formal privacy definition, **independent from prior information**.
- **Easy to compute**, independently of the number of locations.
- No need of a trusted third-party.

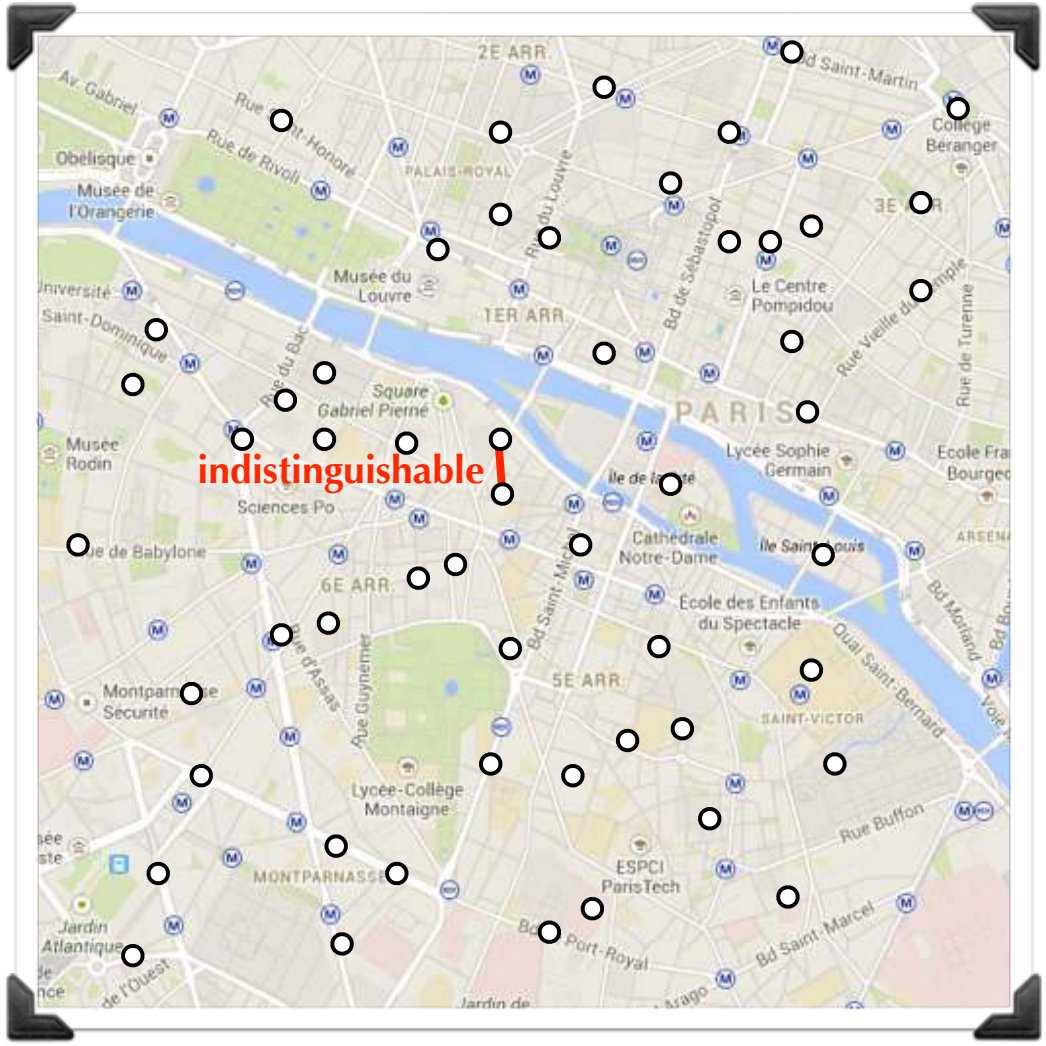
Towards a Definition

- Secrets are **locations**.
- Attacker's goal: **distinguish** location x from x′.
- The closer two locations are, the more indistinguishable they should be.



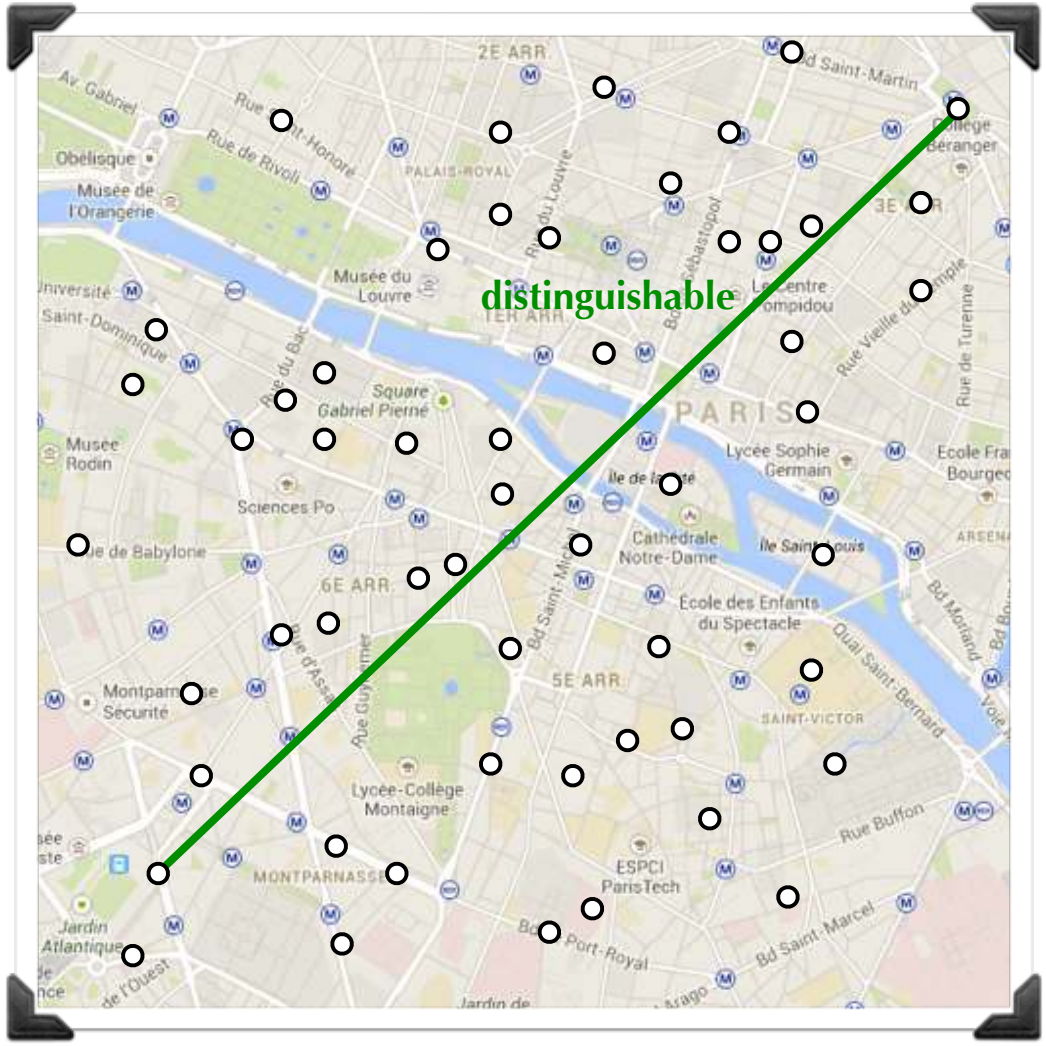
Towards a Definition

- Secrets are **locations**.
- Attacker's goal: **distinguish** location x from x′.
- The closer two locations are, the more indistinguishable they should be.



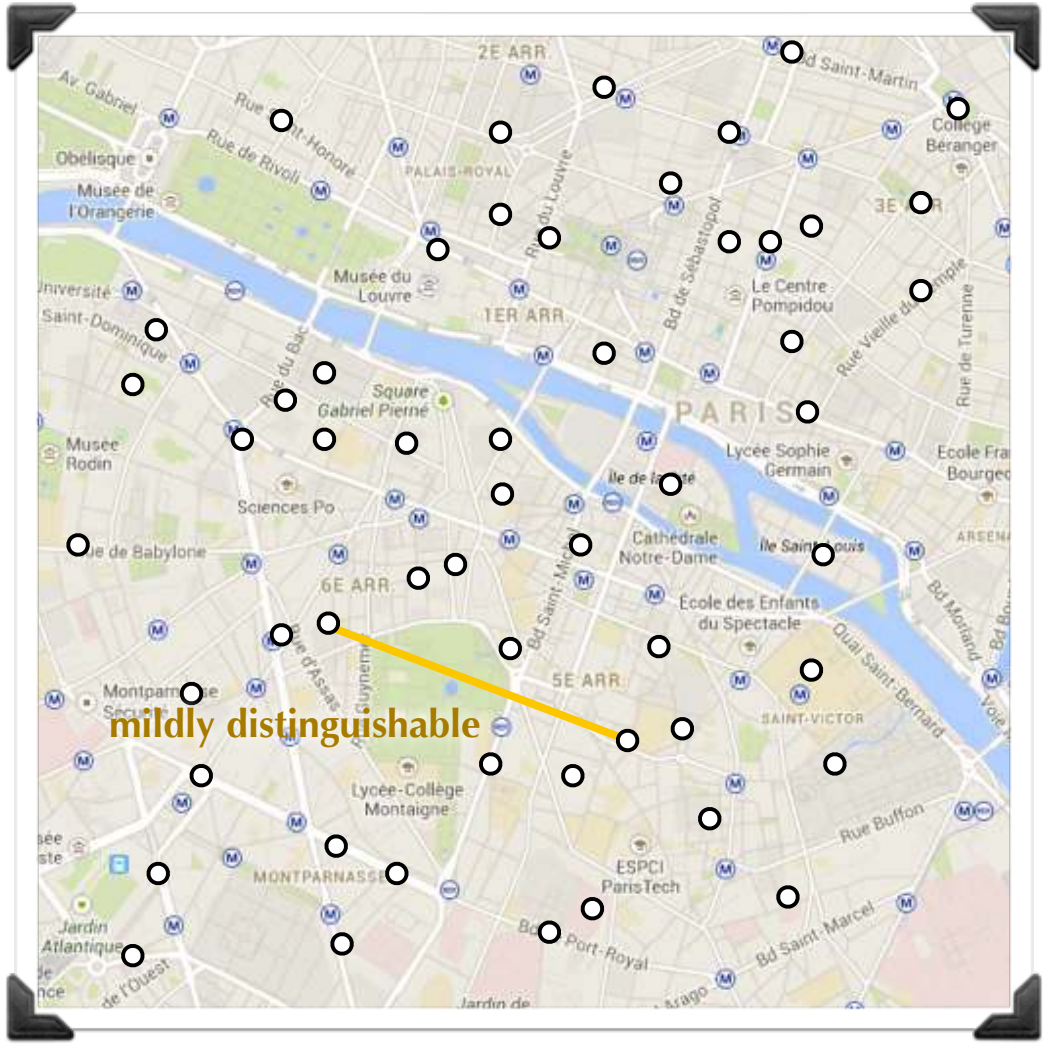
Towards a Definition

- Secrets are **locations**.
- Attacker's goal: **distinguish** location x from x′.
- The closer two locations are, the more indistinguishable they should be.



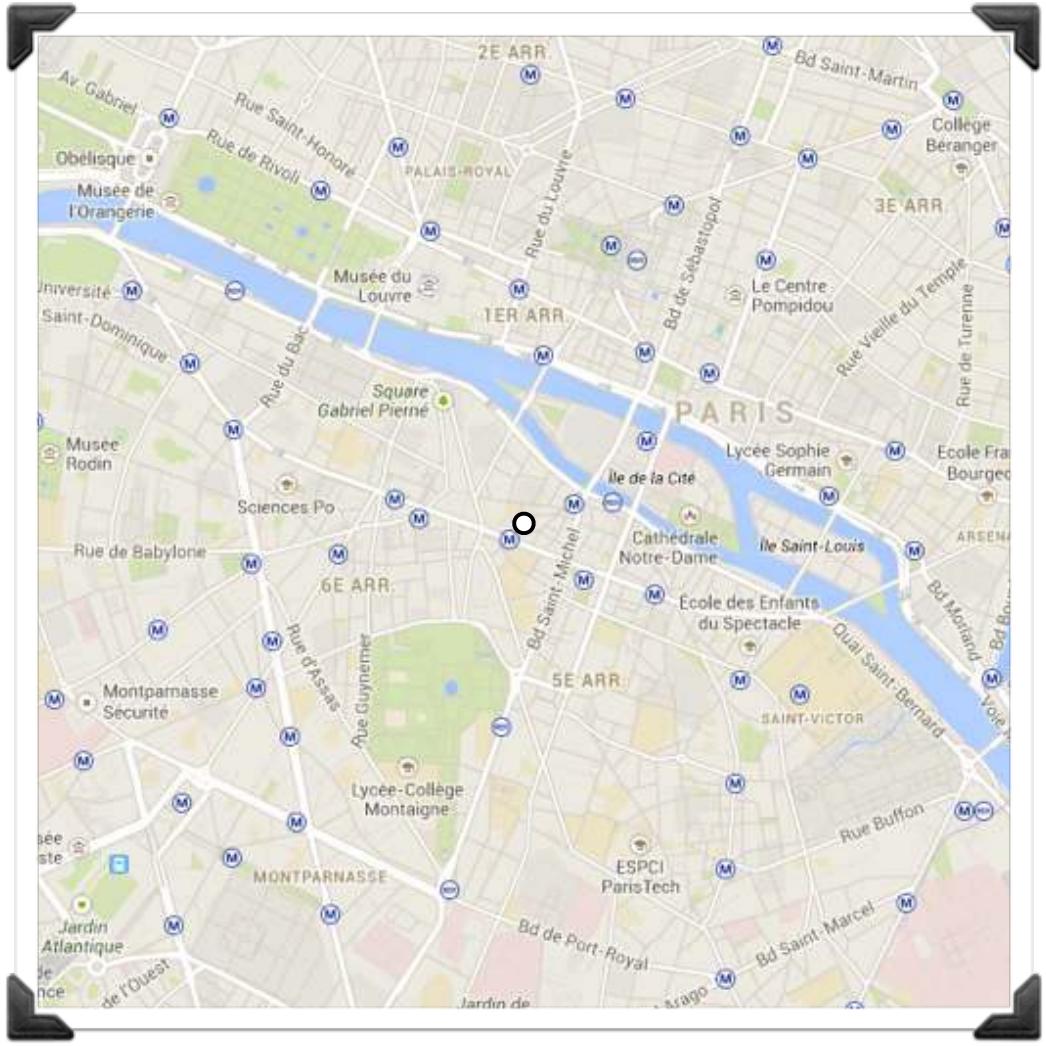
Towards a Definition

- Secrets are **locations**.
- Attacker's goal: **distinguish** location x from x′.
- The closer two locations are, the more indistinguishable they should be.



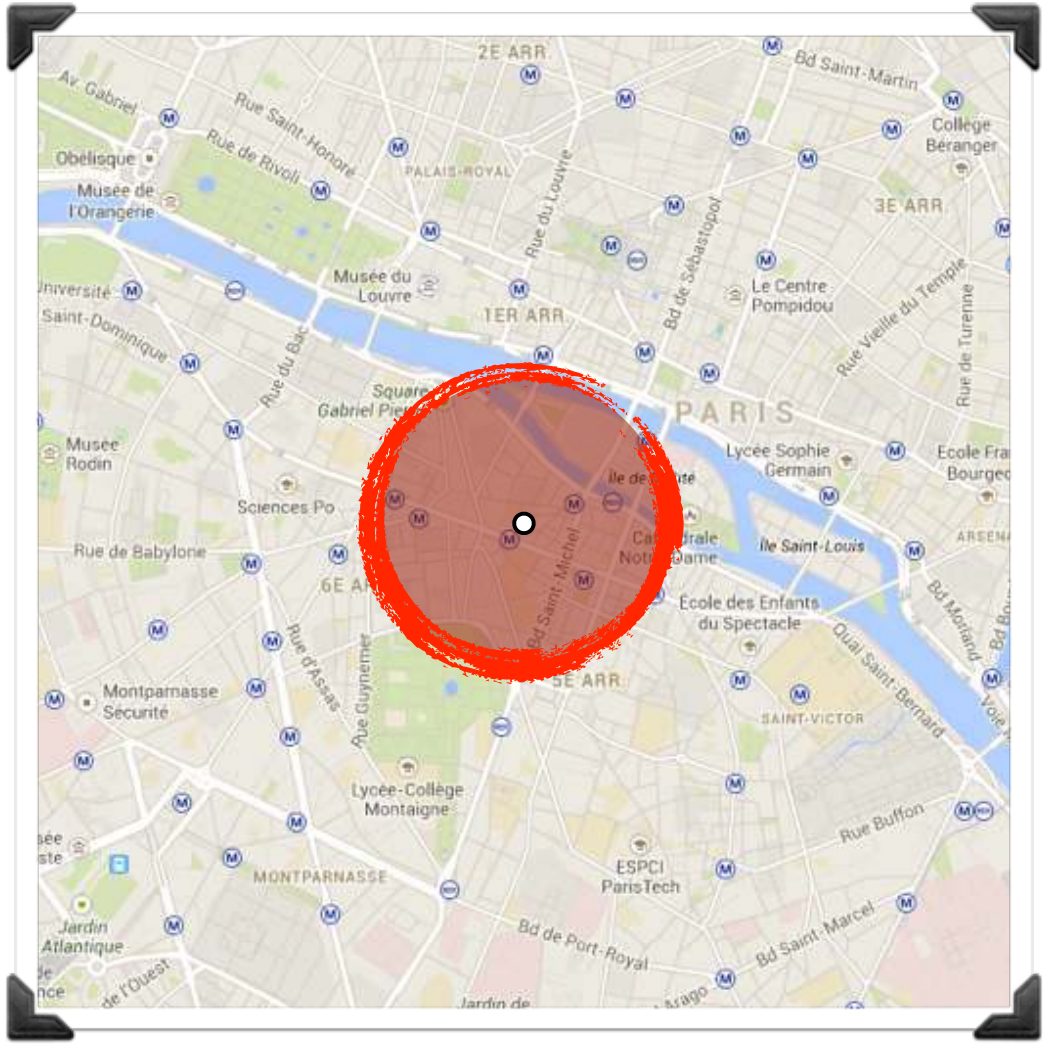
Towards a Definition

- Secrets are **locations**.
- Attacker's goal: **distinguish** location x from x′.
- The closer two locations are, the more indistinguishable they should be.



Towards a Definition

- Secrets are **locations**.
- Attacker's goal: **distinguish** location x from x′.
- The closer two locations are, the more indistinguishable they should be.

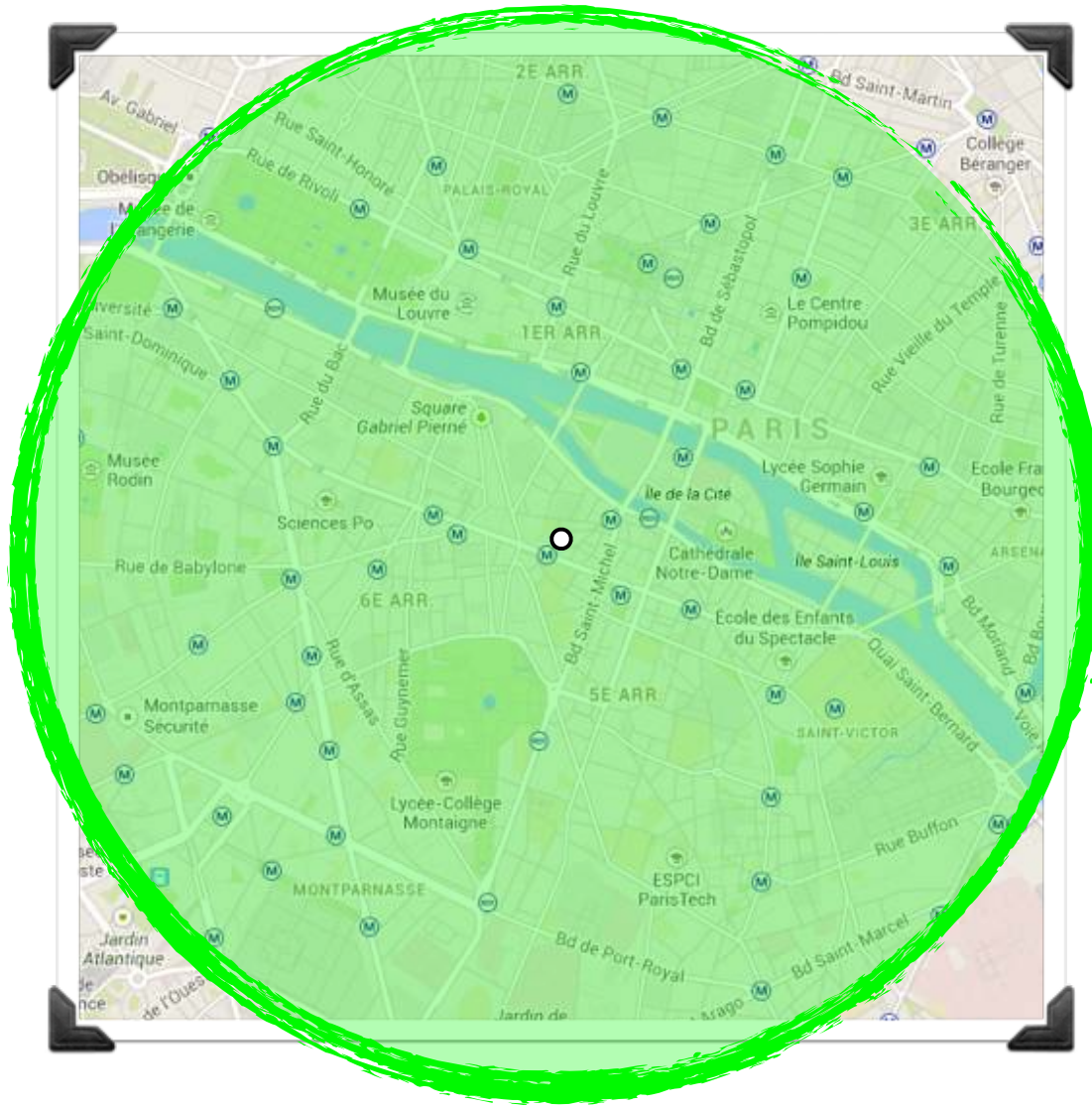


Towards a Definition

- Secrets are **locations**.
- Attacker's goal: **distinguish** location x from x′.
- The closer two locations are, the more indistinguishable they should be.

Towards a Definition

- Secrets are **locations**.
- Attacker's goal: **distinguish** location x from x′.
- The closer two locations are, the more indistinguishable they should be.



Geo-Indistinguishability

- We can consider the **set of possible locations** as the set of secrets, and the **Euclidian distance** as the metric.

A location obfuscation mechanism M provides ϵ-geo-indistinguishability if:

$$\mathcal{D}_p(M(x), M(x')) \leq \epsilon\, d(x,x') \qquad \forall\ x, x'$$

Where d(x,x′) is the Euclidean distance between x and x′.

[Pierce et al., ICFP 2010]

[Chatzikokolakis et al, PETS 2013]

Line of work

[PETS'13] privacy under general metrics

[CCS'13] application to location privacy, planar Laplace

[CCS'14] mechanisms of optimal utility

[PETS'14] protecting location traces

[ongoing] privacy metrics adapted to the semantics of the map

The Planar Laplace Mechanism

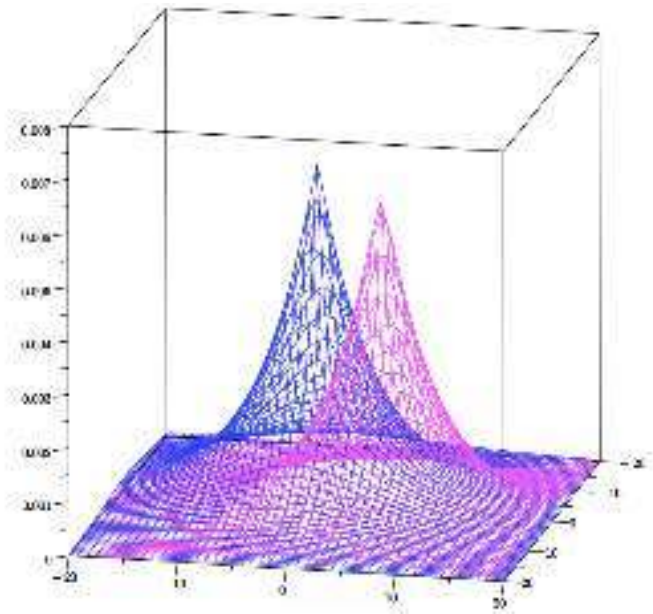
A way to achieve geo-indistinguishability is to add noise from a 2-dimensional Laplace distribution.

Computationally efficient.

Scales very well.

Independent from the set of locations and the user.

Utility may not be optimal.



Utility of a mechanism

We measure the (inverse of) utili

π : user's prior
d_Q : quality metric

Utility measure:

$\mathrm{QL}(K) =$ **Expected distance of K (wrt π and d_Q)**

Utility depends on the user!

Goal

Guarantee geo-indistinguishability.

- Pre-fixed privacy level ϵ.
- Independent from the user and adversary's prior.

Optimize utility.

- For a given set of locations.
- Depends on the user's prior π.

The $d_{\mathcal{X}}$-optimal mechanism

K is OptQL wrt ϵ, π, $d_{\mathcal{X}}$ and d_Q iff:

From all mechanisms that provide geo-indistinguishability with level at least ϵ, K is the one with the best utility.

The $d_{\mathcal{X}}$-optimal mechanism

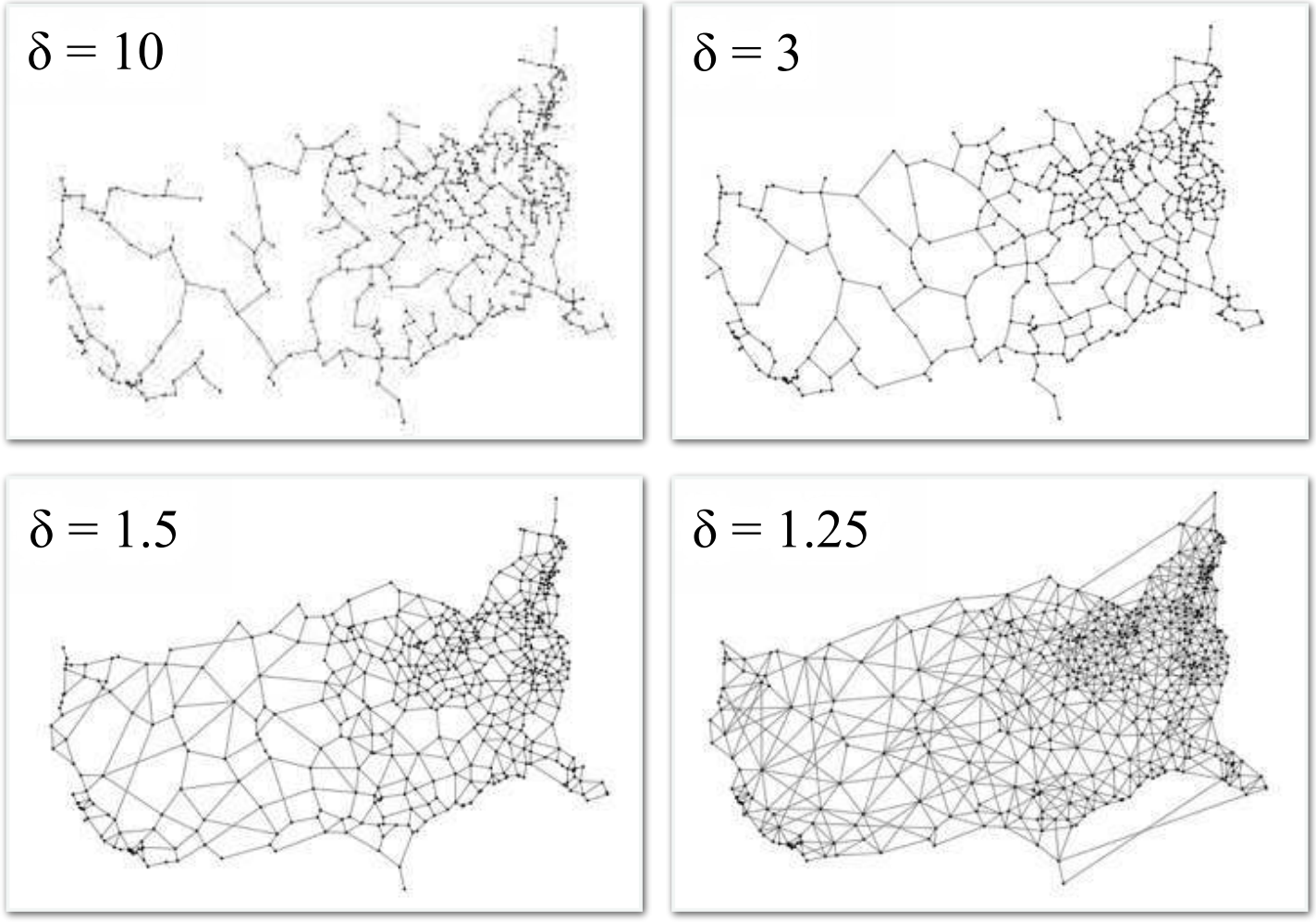
We get K by solving a linear optimization problem:

Choose:	K		
To minimize:	$\mathrm{QL}(K)$		
Subject to:	$k_{xz} \leq e^{\epsilon d_{\mathcal{X}}(x,x')} k_{x'z}$	$\forall\, x,x',z$	($d_{\mathcal{X}}$-privacy)

$|\mathcal{X}|^3$ constraints!

Because we need to consider the privacy constraints for all x, x'.

Spanners



Images from “Geometric Spanner Networks”, by G. Narasimhan and M. Smid

Protecting location traces

- Secrets are now tuples $\mathbf{x} = (x_1, \ldots, x_n)$
- Distance between tuples:

$$d_\infty(\mathbf{x}, \mathbf{x}') = \max_i d(x_i, x_i')$$

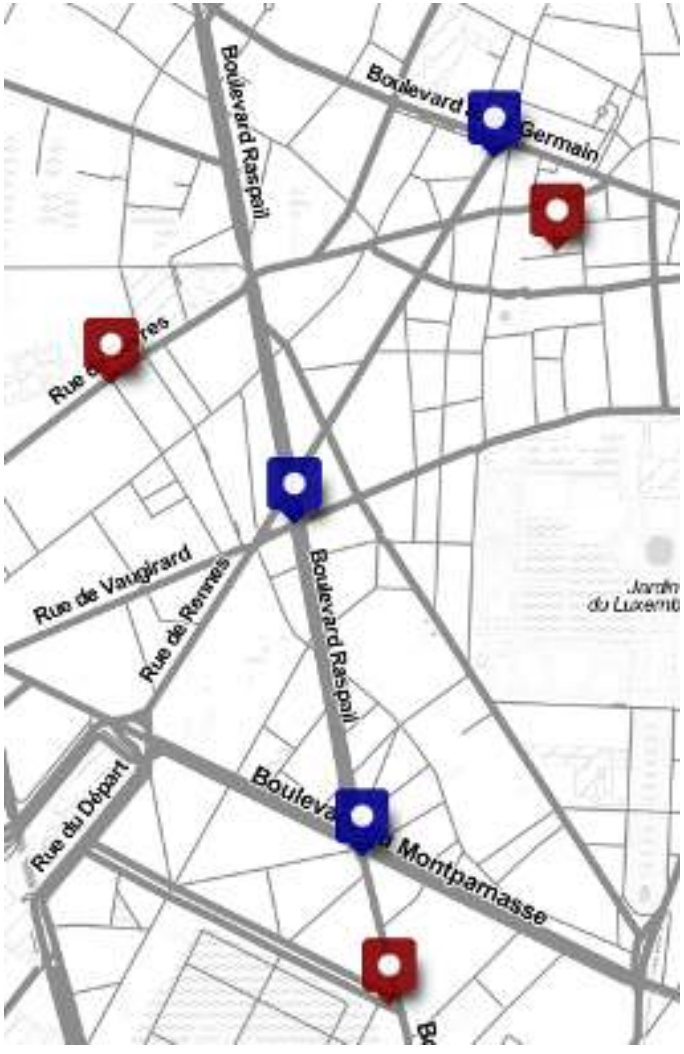
- Use ϵd_∞-privacy

Independent Mechanism

apply noise to each point

$n \epsilon_N$ d_∞-private

- works on any trace (including random teleporting)
- budget is linear on n



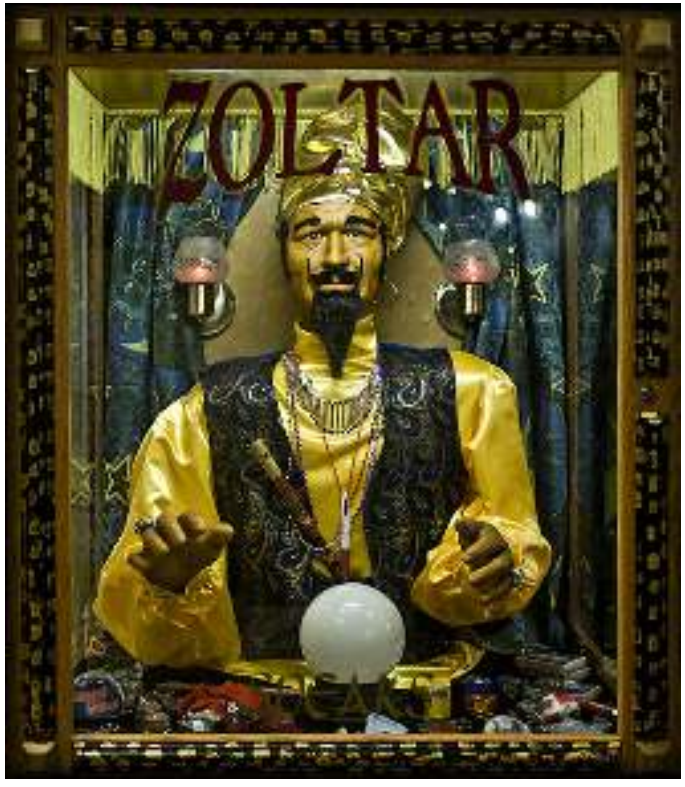
Predictive Mechanism

prediction function

- based on public info
- obtain point $\tilde{z}_i$

is $\tilde{z}_i$ close to x_i?

- yes: report $\tilde{z}_i$
- no: add new noise to x_i



Predictive Mechanism

prediction function

- based on public info
- obtain point $\tilde{z}_i$

is $\tilde{z}_i$ close to x_i?

- yes: report $\tilde{z}_i$
- no: add new noise to x_i

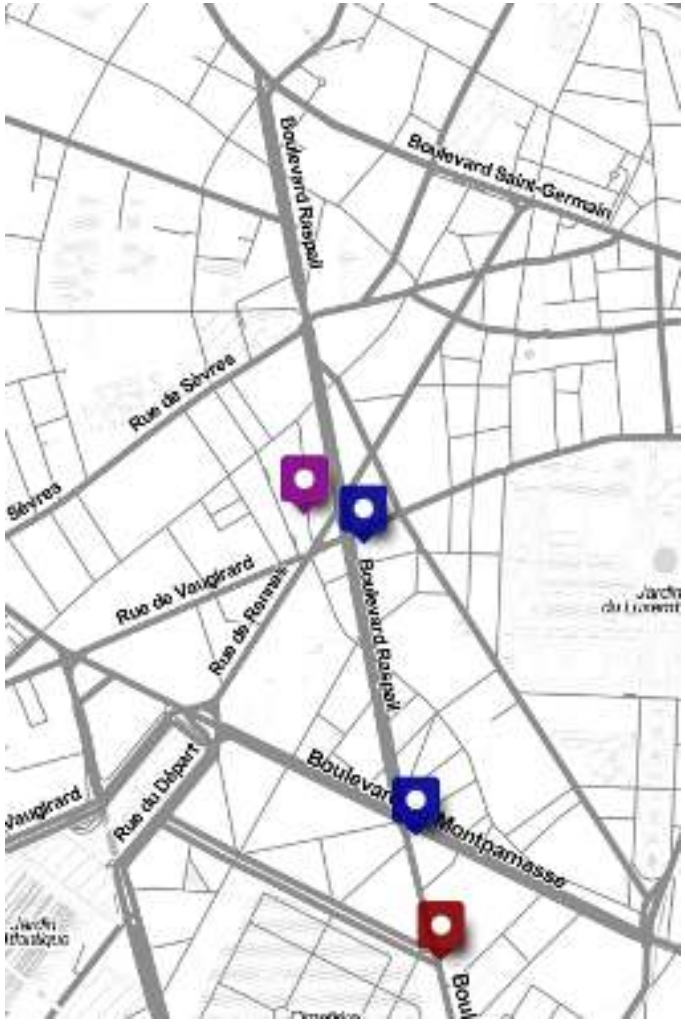
Predictive Mechanism

prediction function

- based on public info
- obtain point $\tilde{z}_i$

is $\tilde{z}_i$ close to x_i?

- yes: report $\tilde{z}_i$
- no: add new noise to x_i



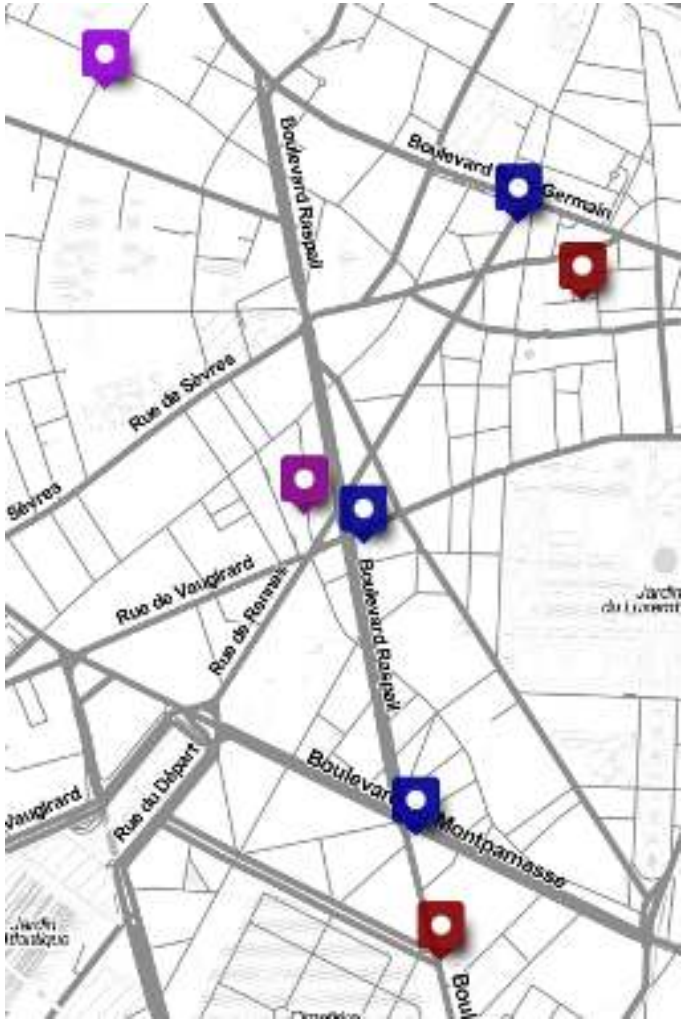
Predictive Mechanism

prediction function

- based on public info
- obtain point $\tilde{z}_i$

is $\tilde{z}_i$ close to x_i?

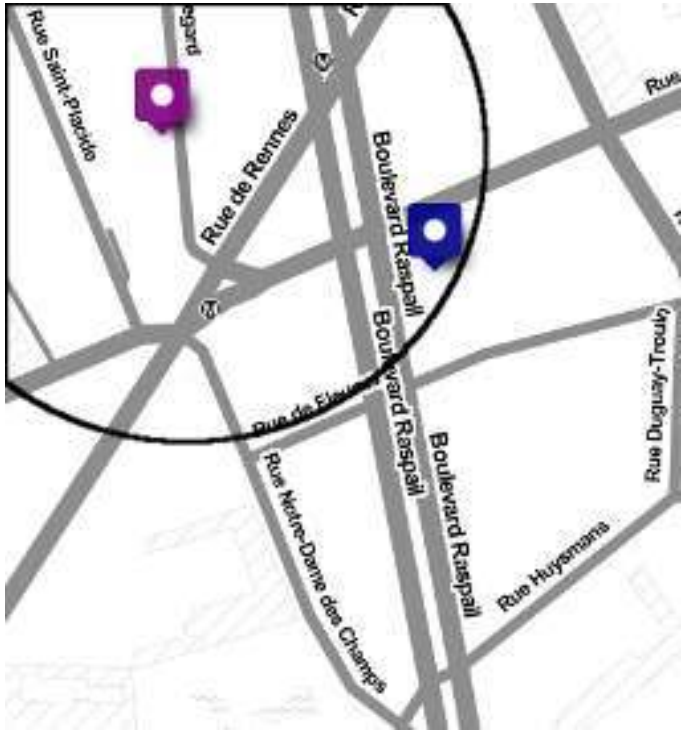
- yes: report $\tilde{z}_i$
- no: add new noise to x_i



Testing the prediction

Deterministic test

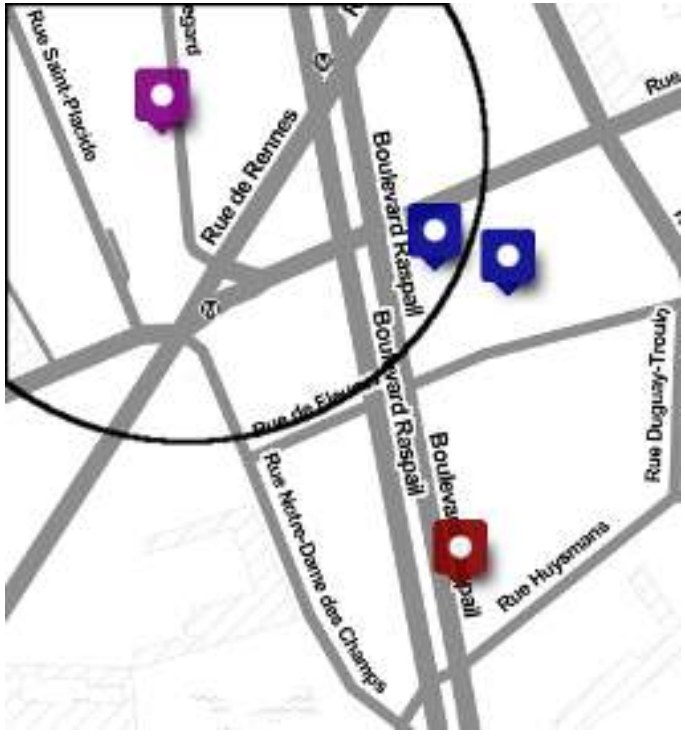
breaks privacy



Testing the prediction

Deterministic test

breaks privacy



Testing the prediction

Deterministic test

breaks privacy

D-Private test

use a noisy border for the test

Testing the prediction

Deterministic test

breaks privacy

D-Private test

use a noisy border for the test

Budget used at each step

ϵ_θ (successful prediction)
or $\epsilon_\theta + \epsilon_N$ (new noise)

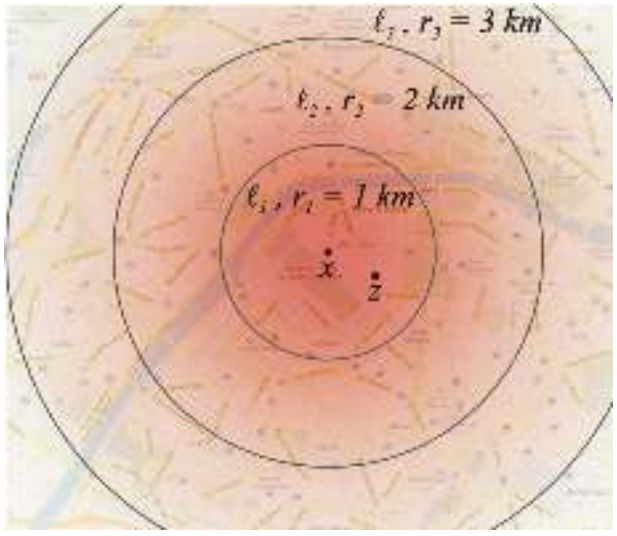
(In)Distinguishability Metric

What is it that you want to be similar to?

(and how much?)

Euclidean Metric

- space provides privacy
- scaled by ϵ

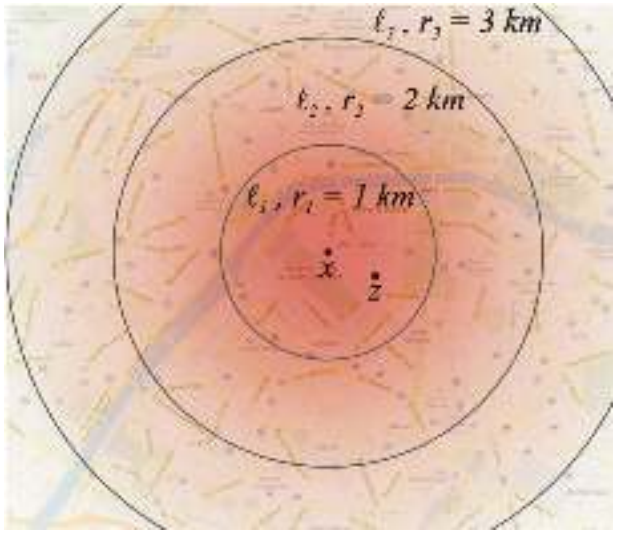


Euclidean Metric

- space provides privacy
- scaled by ϵ

but...

- space is not equally valuable everywhere
- POI/population/... also provide privacy
- we can achieve better privacy/utility by adapting the noise to the map



Building a custom metric

- divide the space in cells (eg grid 100m x 100m)
- privacy weight of each cell
  - from POI/population/... (eg by querying OSM)
  - from the cell's area
- build a metric d satisfying the requirement f:

$$\text{weight}(B_r^d(x)) \geq f(r) \qquad x, r$$

Building a custom metric

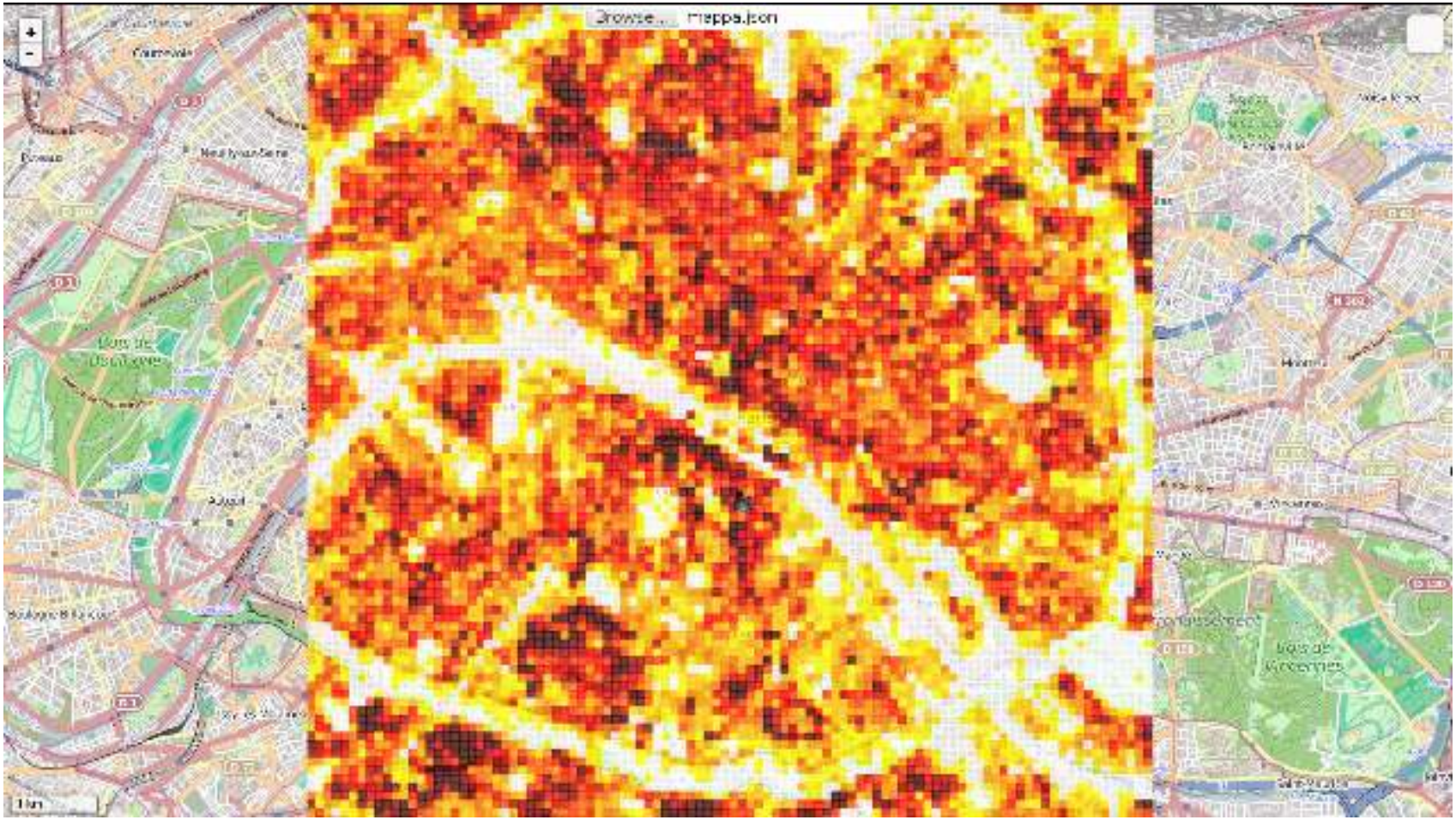
- divide the space in cells (eg grid 100m x 100m)
- privacy weight of each cell
  - from POI/population/... (eg by querying OSM)
  - from the cell's area
- build a metric d satisfying the requirement f:

$$\text{weight}(B_r^d(x)) \geq f(r) \qquad x, r$$

Exponential Mechanism

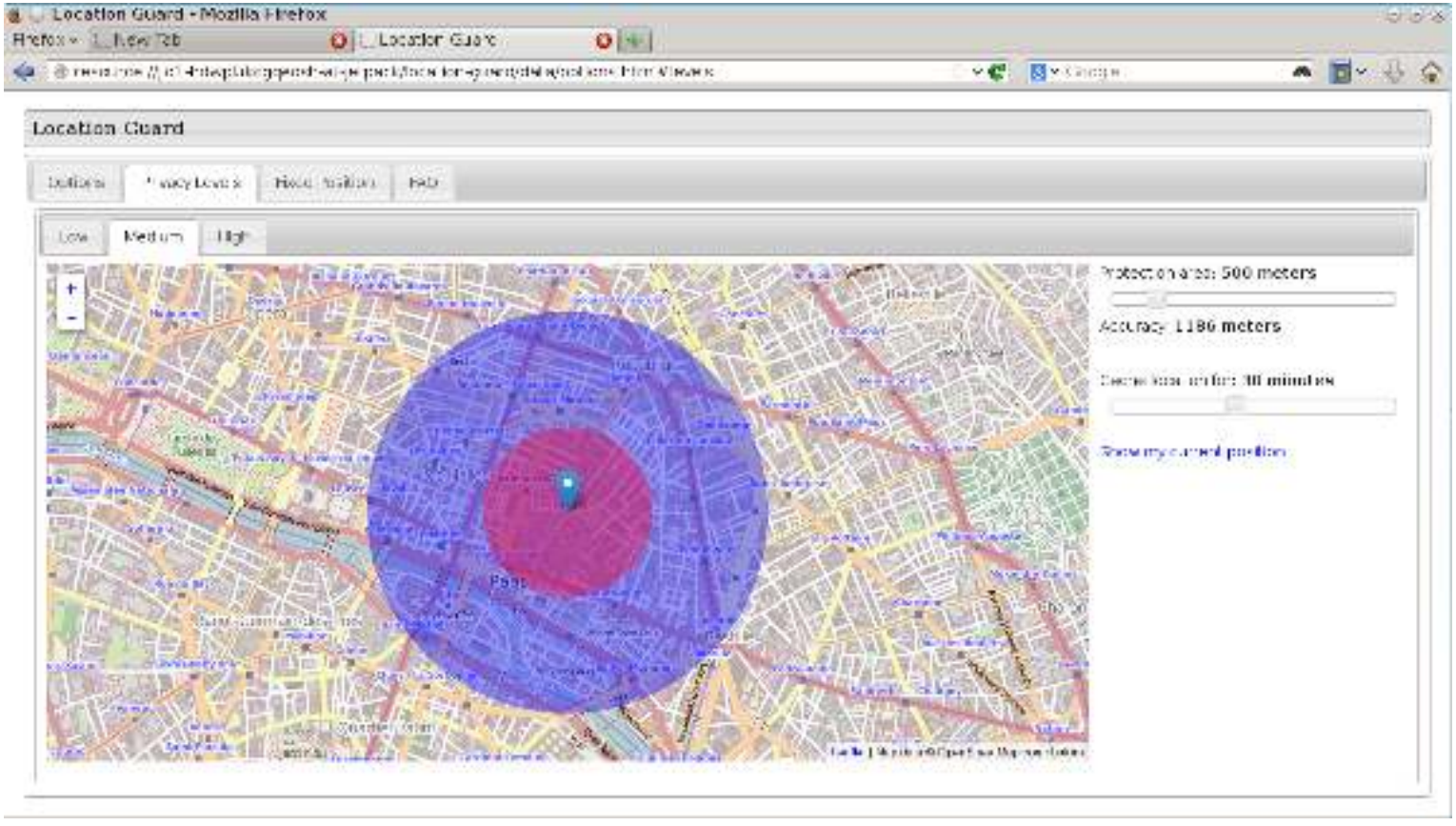
constructed from any metric d

Privacy weights



Obtained Mechanism

Location Guard for Chrome and Firefox



`https://github.com/chatziko/location-guard`

4700+ daily users

Future work

Privacy guarantees under (un)correlation conditions between the points in the trace.

Questions?