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

Journées Recherche 2016, Labex DigiCosme
Apr 12, 2016

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.

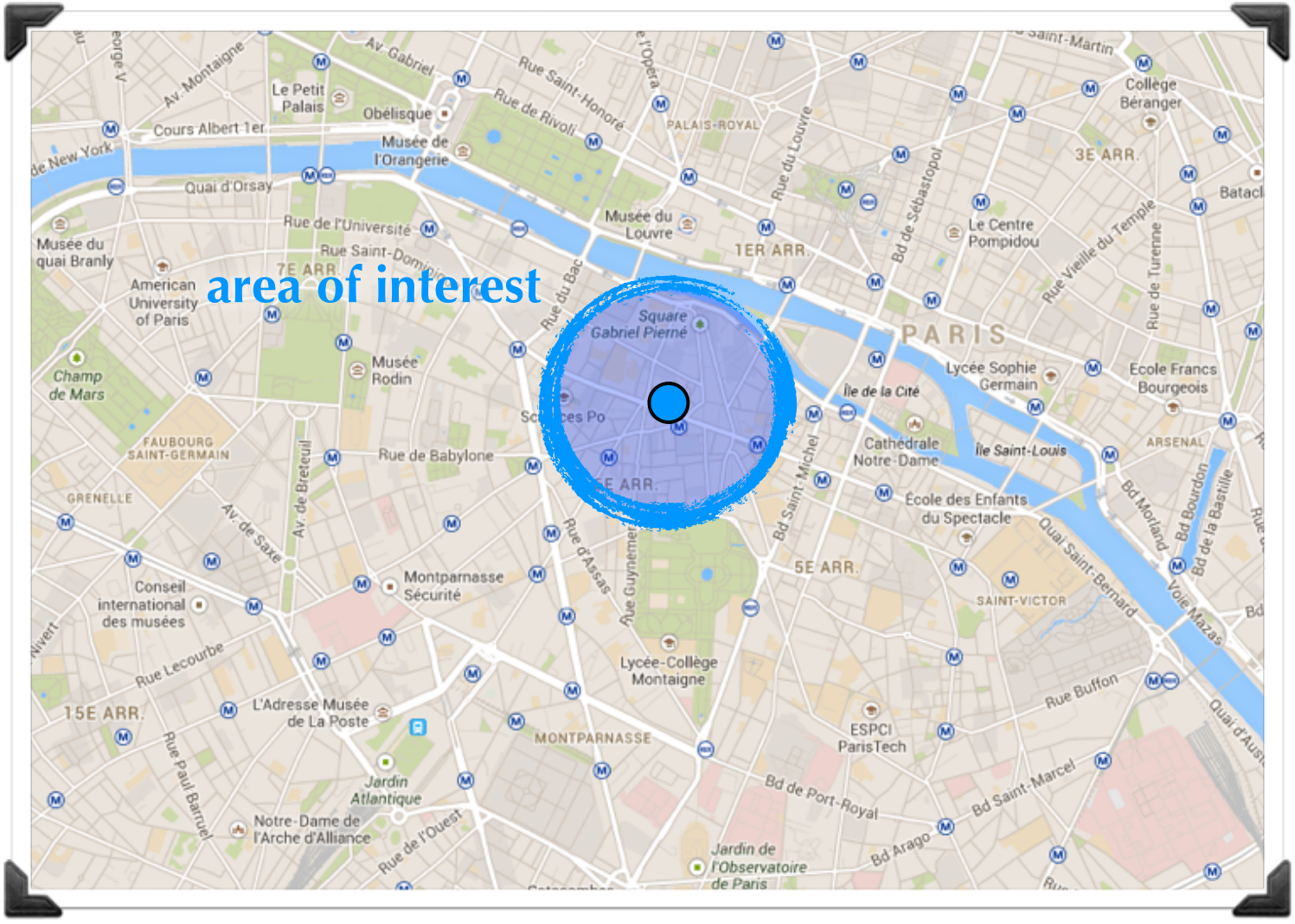
Location-Based Systems

- 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**.

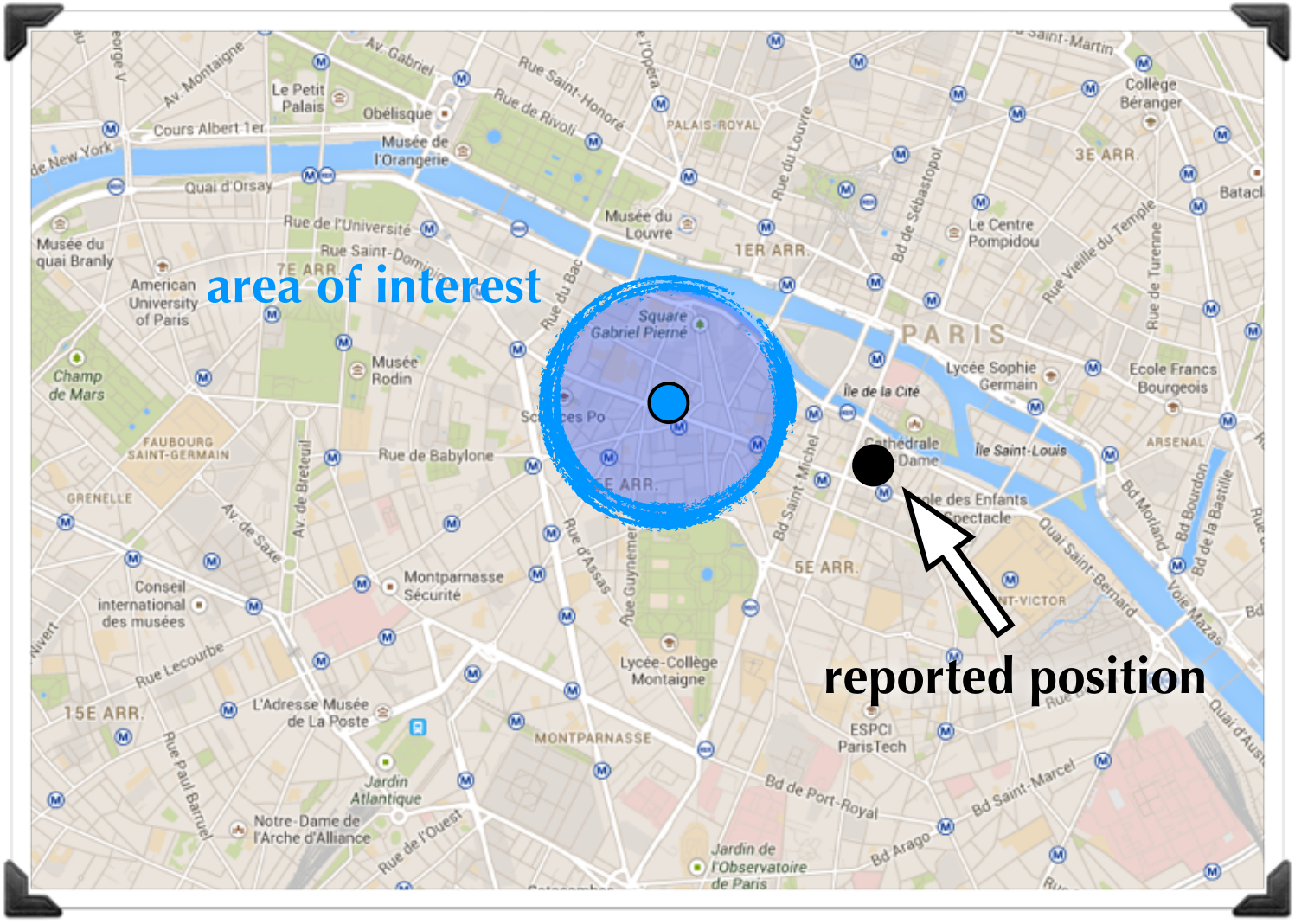
Example

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

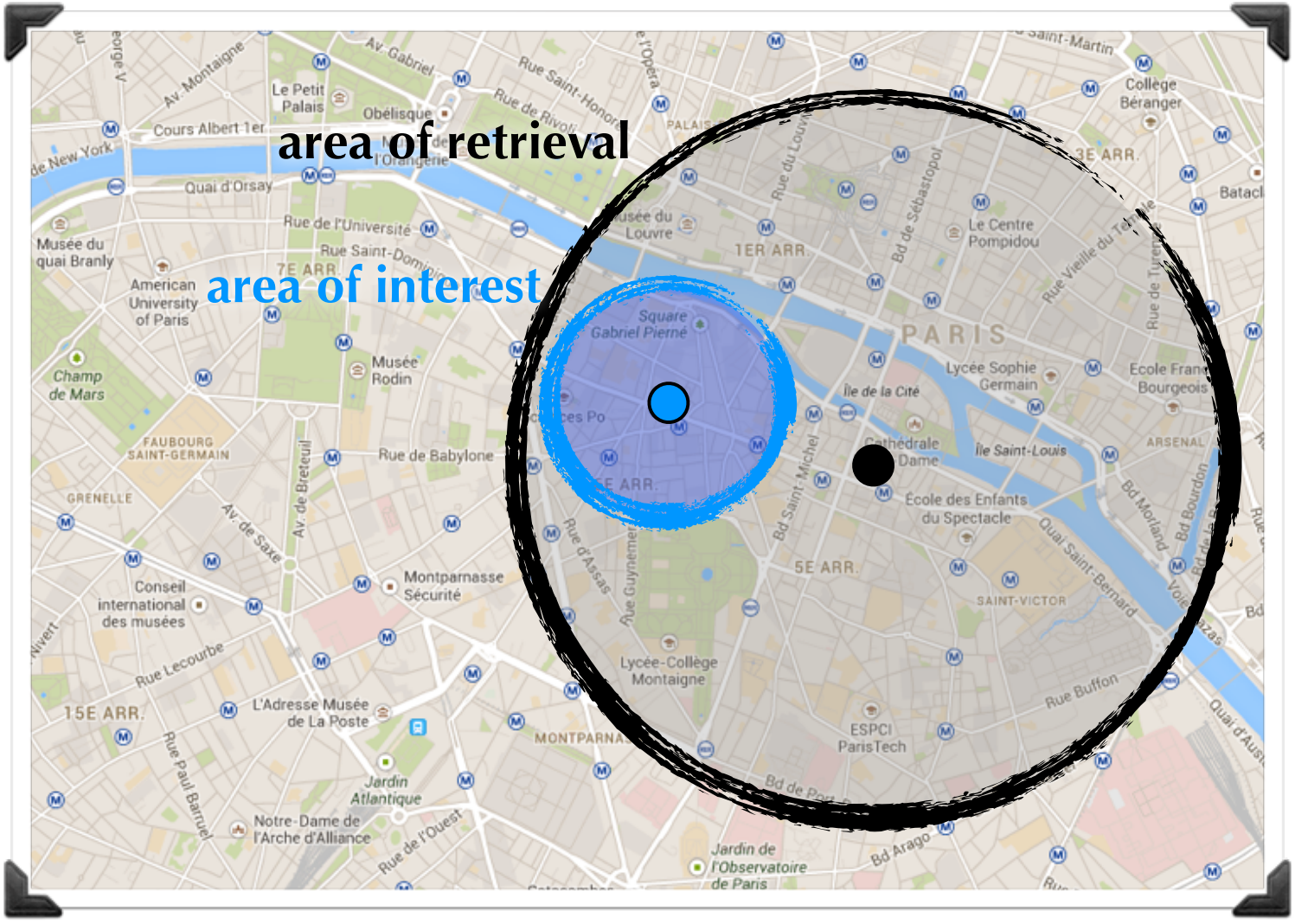
Obfuscation



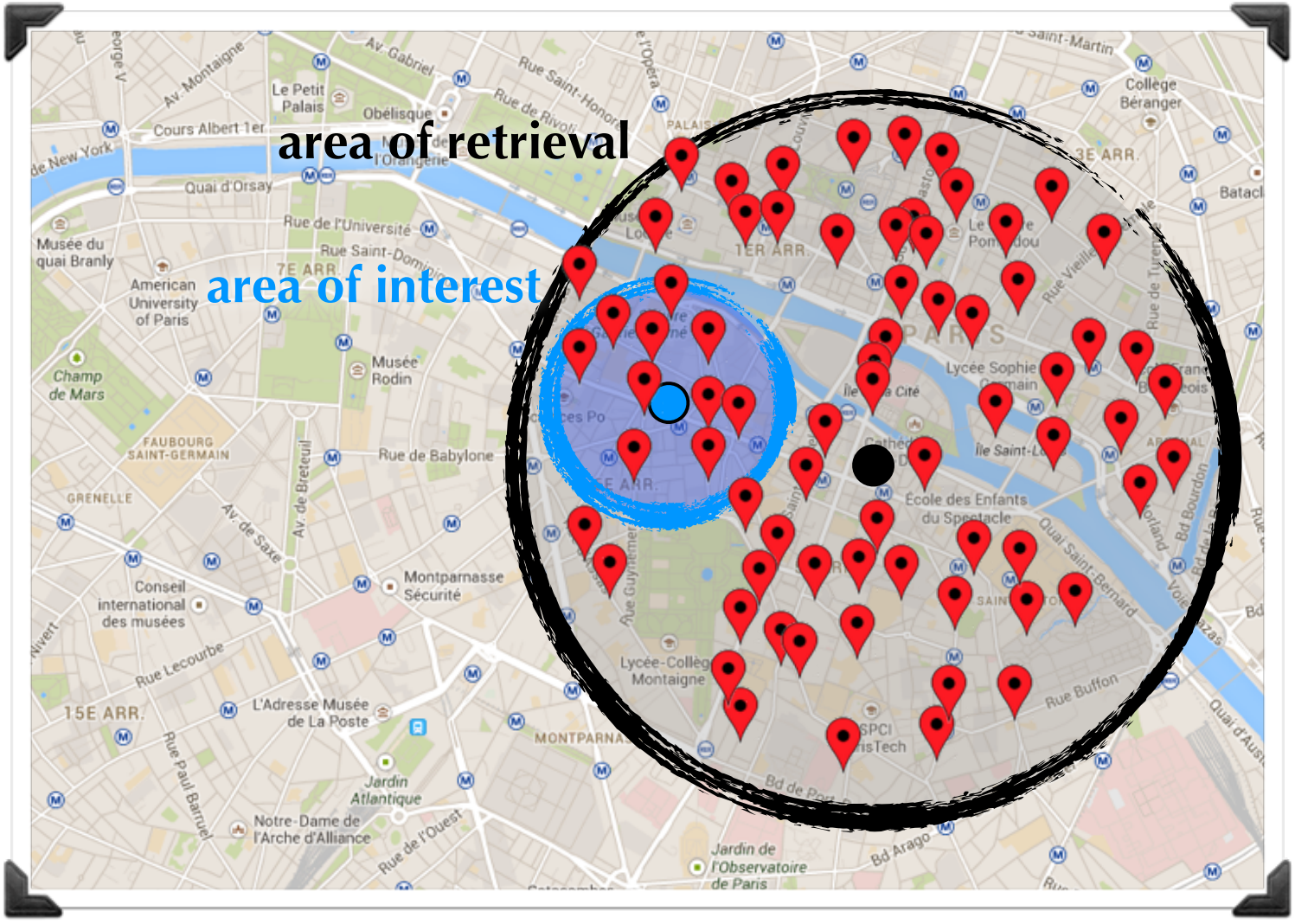
Obfuscation



Obfuscation



Obfuscation



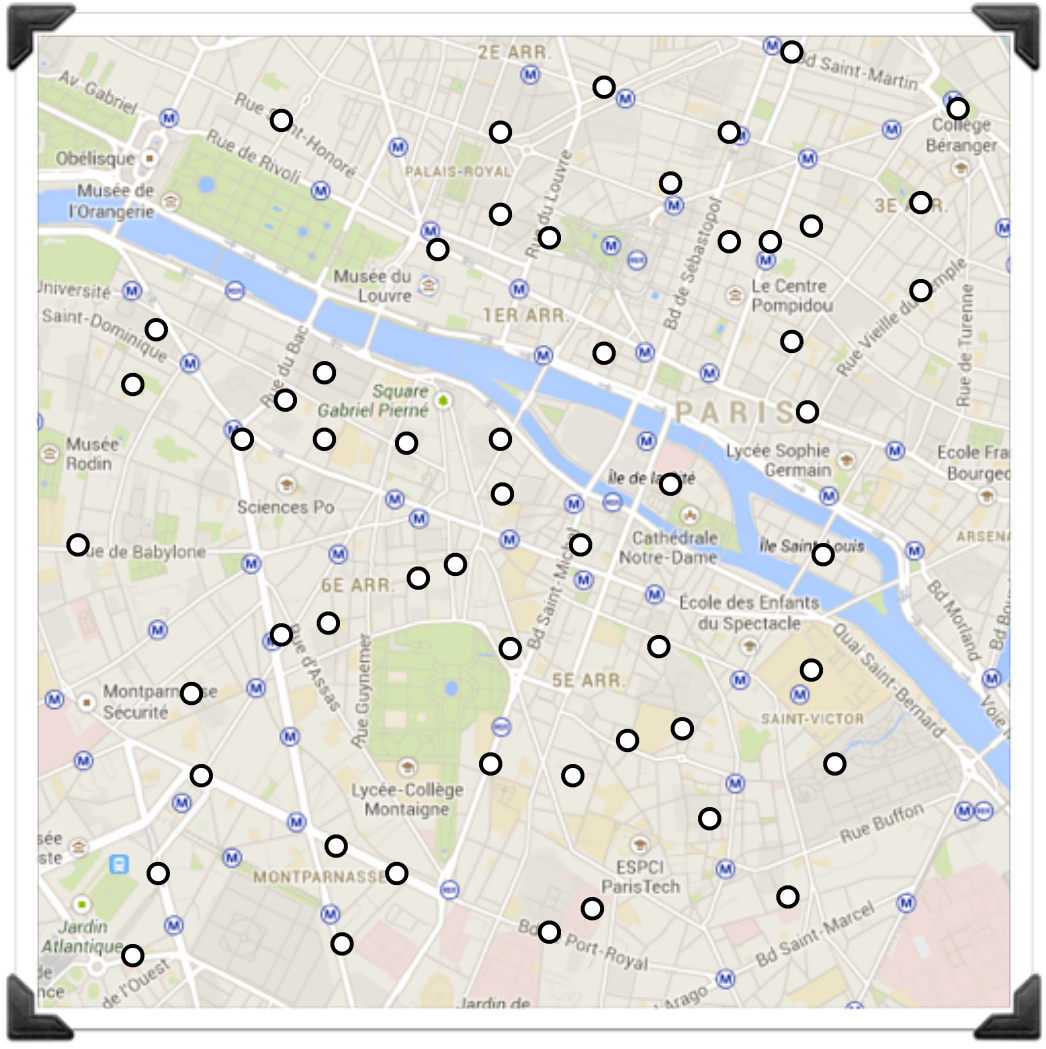
Obfuscation

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.

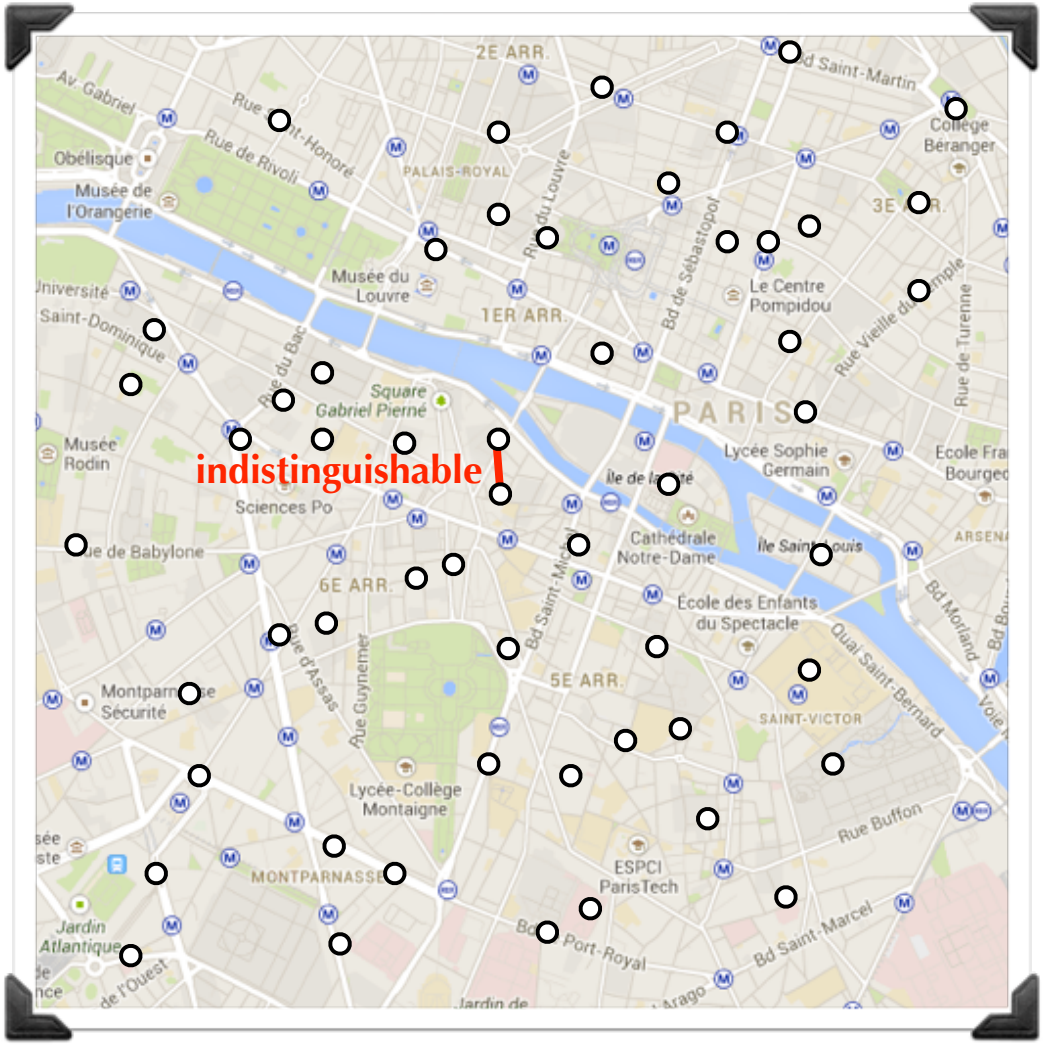
Towards a Definition

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



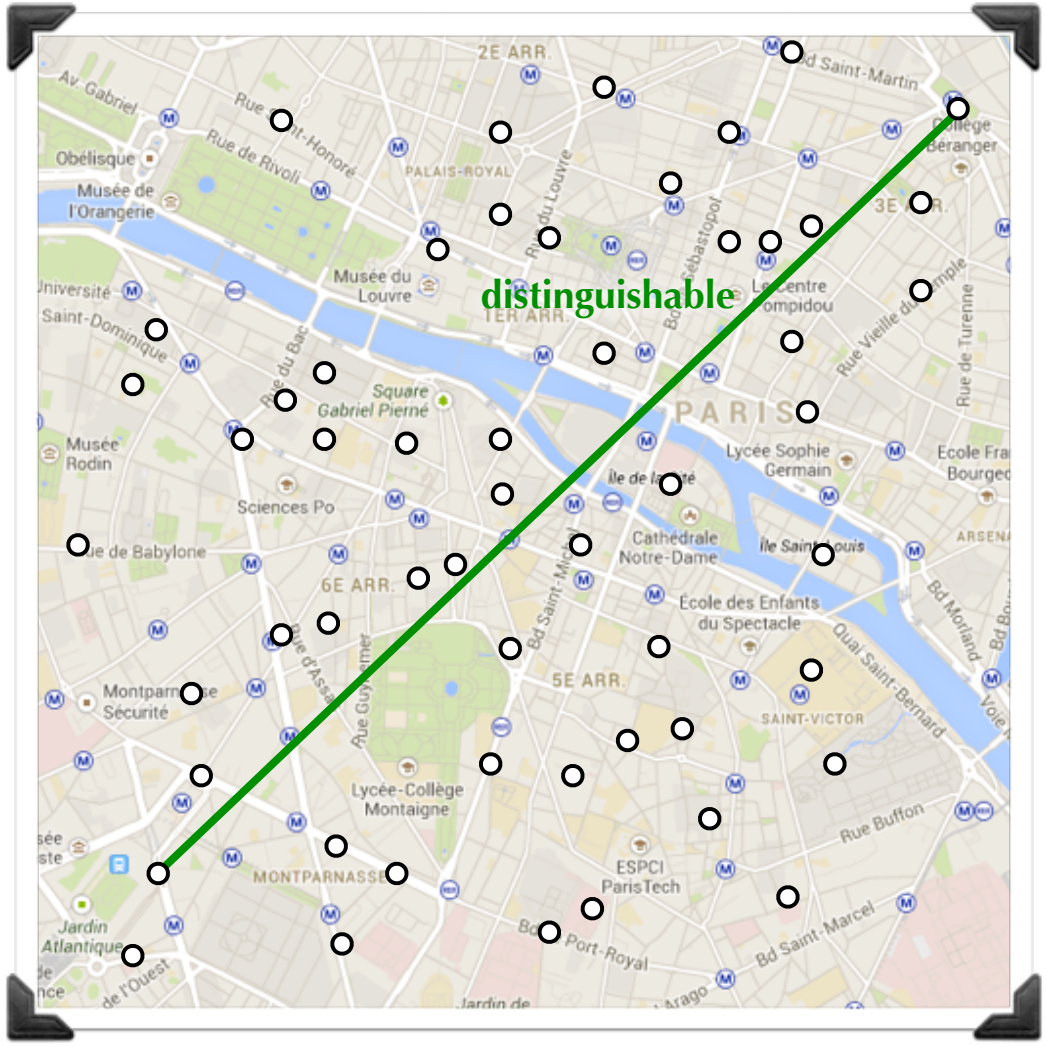
Towards a Definition

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



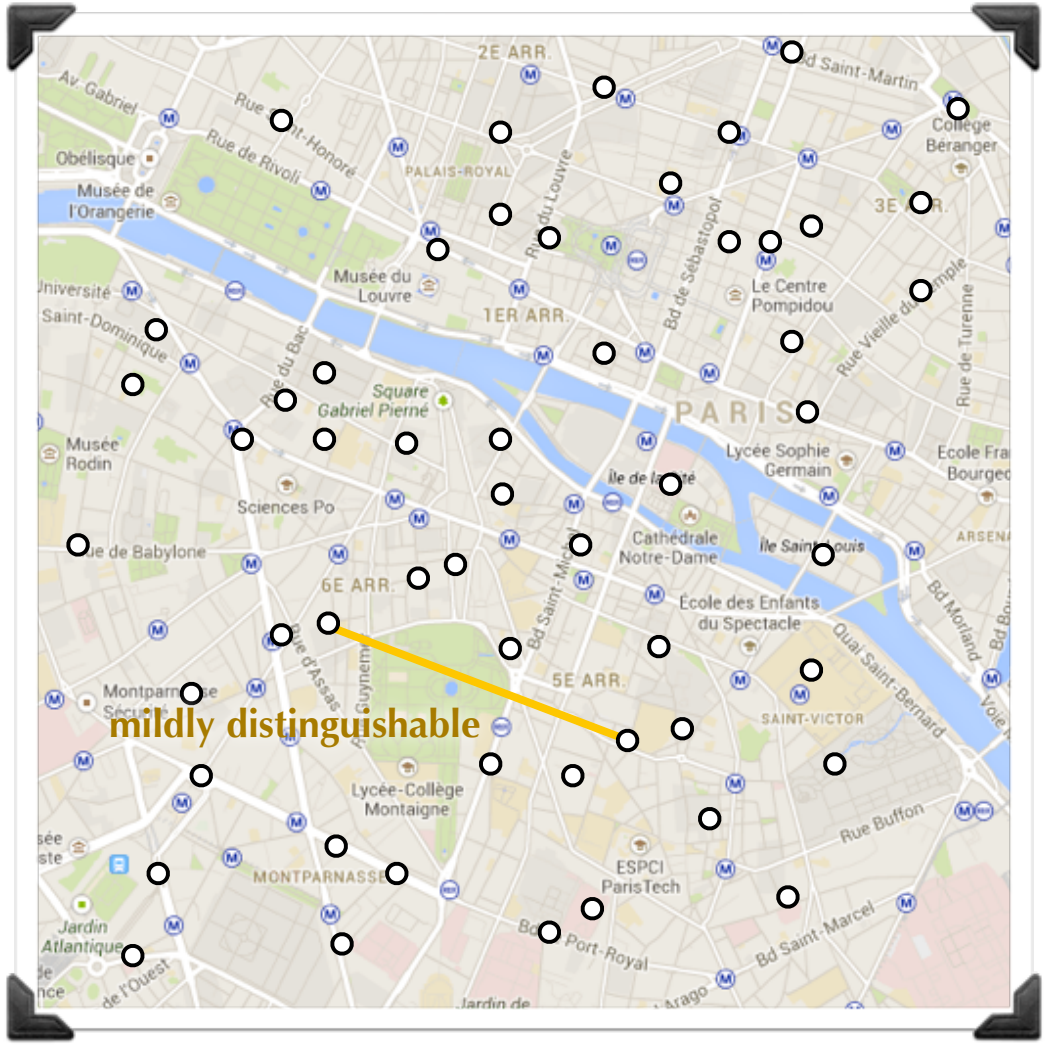
Towards a Definition

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



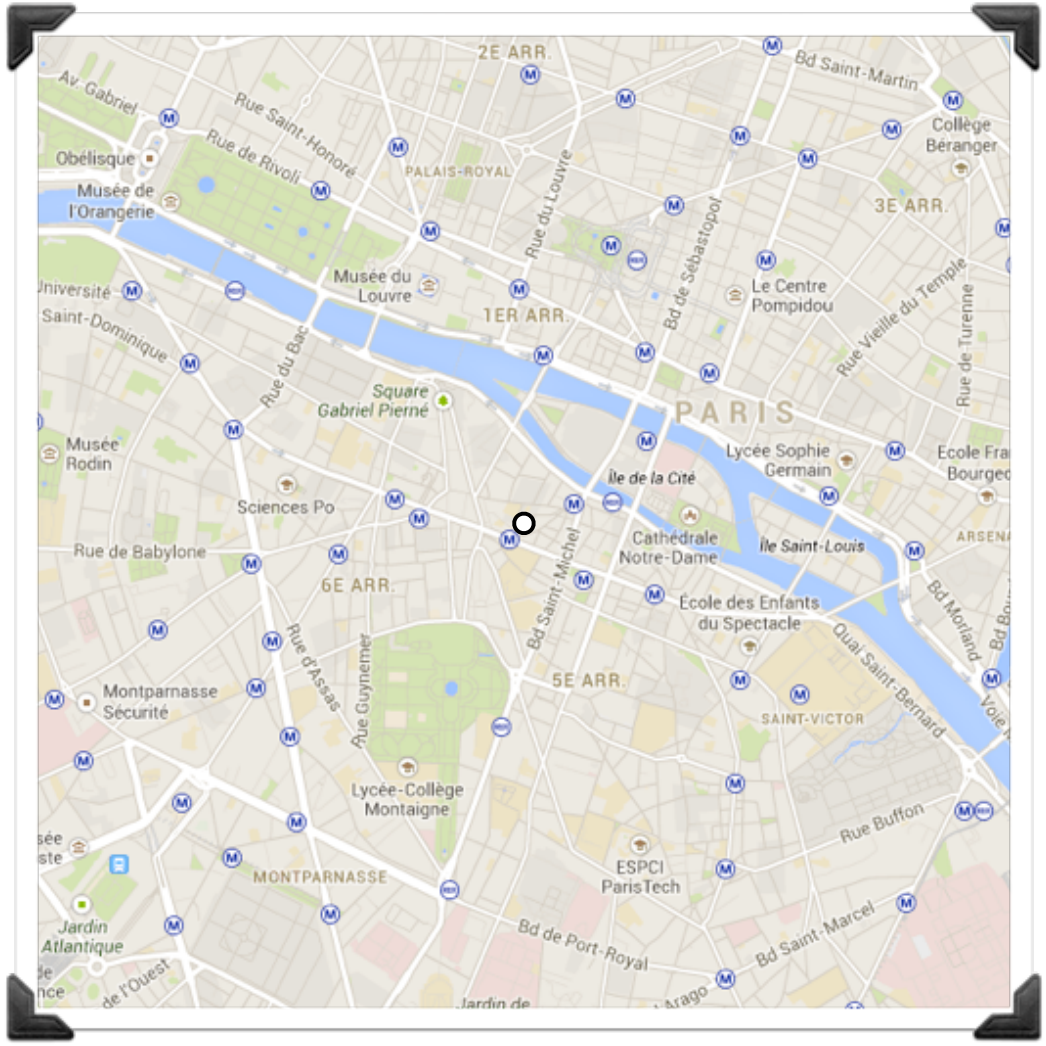
Towards a Definition

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



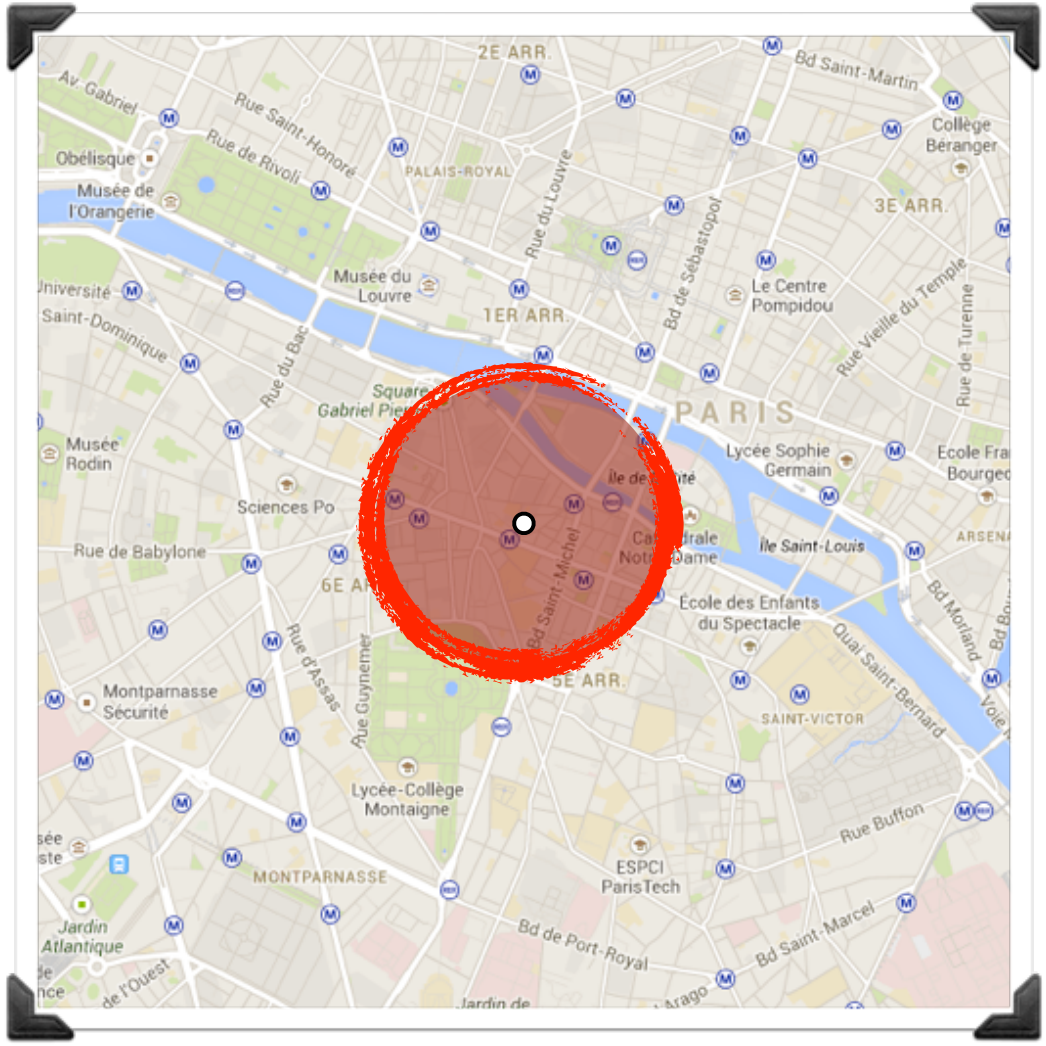
Towards a Definition

- 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.

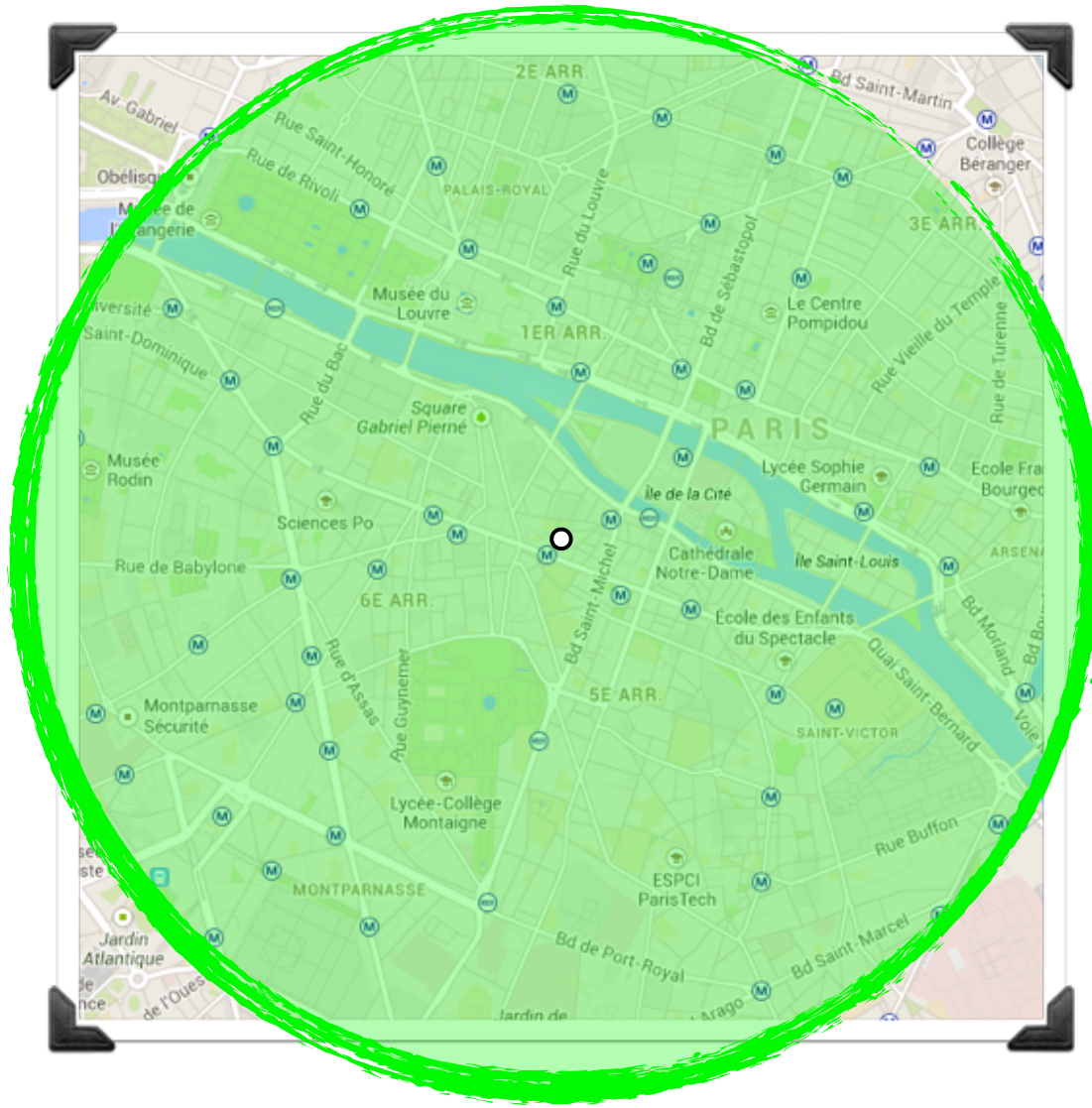


Towards a Definition

- 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

[PETS'15] metrics adapted to the semantics of the map

Tool: Location Guard

Theses:

2014 Nicolás Bordenabe (SIGSAC Award, Prix Polytechnique)

2015 Marco Stronati

The Planar Laplace Mechanism

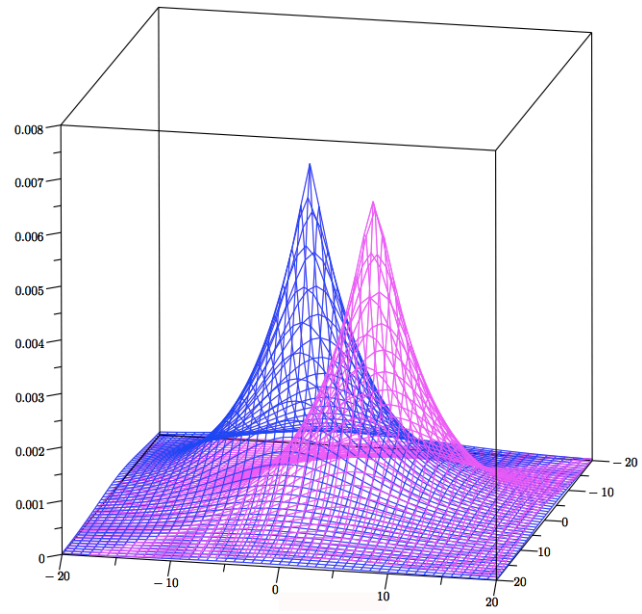
A way to achieve geo-indistinguishability is to add noise from a 2-dimensional Laplace distribution.

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 $d_{\mathcal{Q}}$ 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

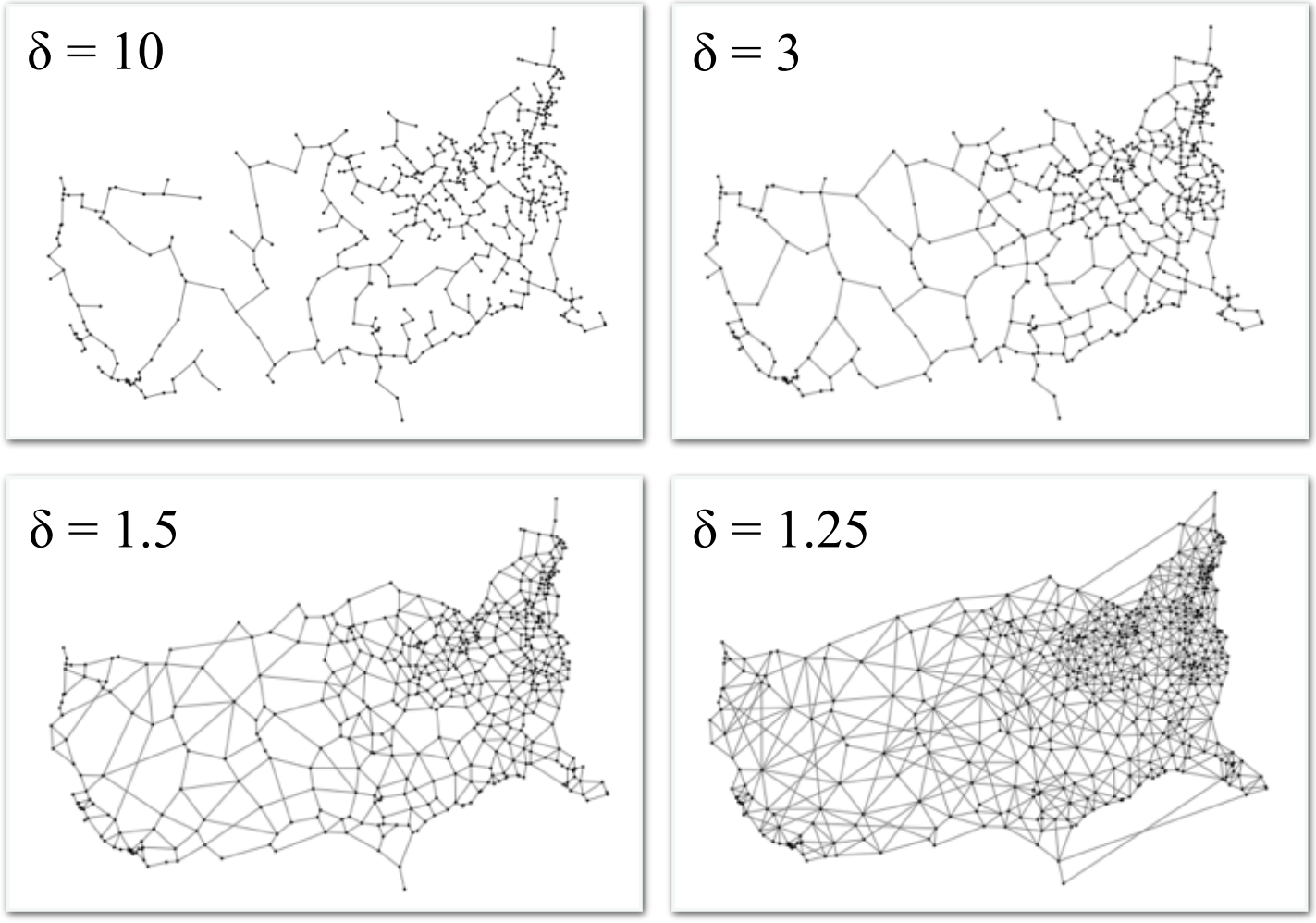
We get K by solving a linear optimization problem:

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

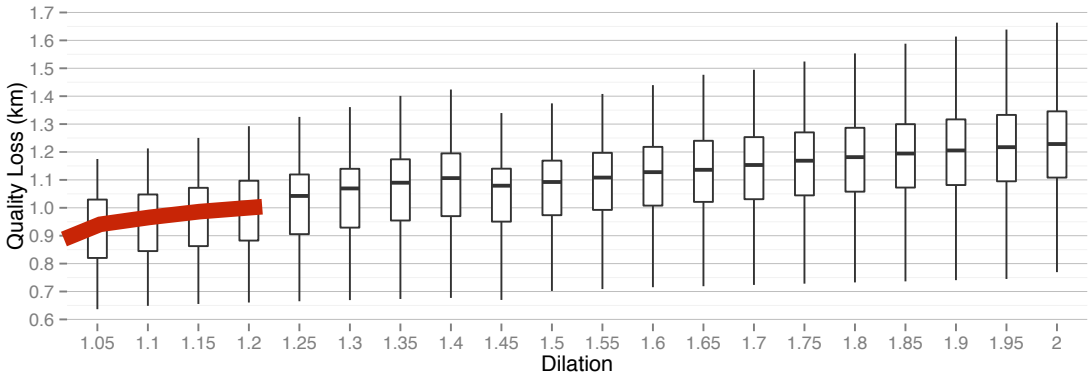


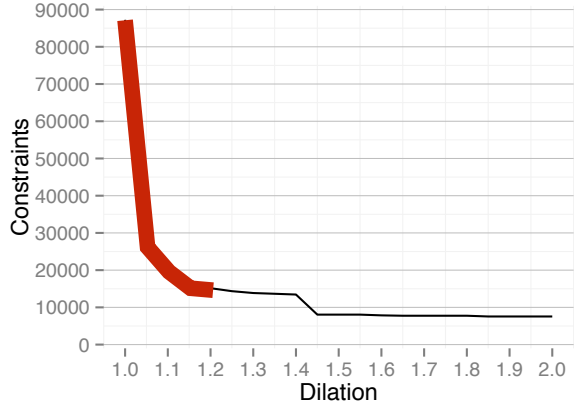
Images from “Geometric Spanner Networks”, by G. Narasimhan and M. Smid

Evaluation

Evaluating the approximation method:

- Effect on the QL
- Reduction in the number of constraints.



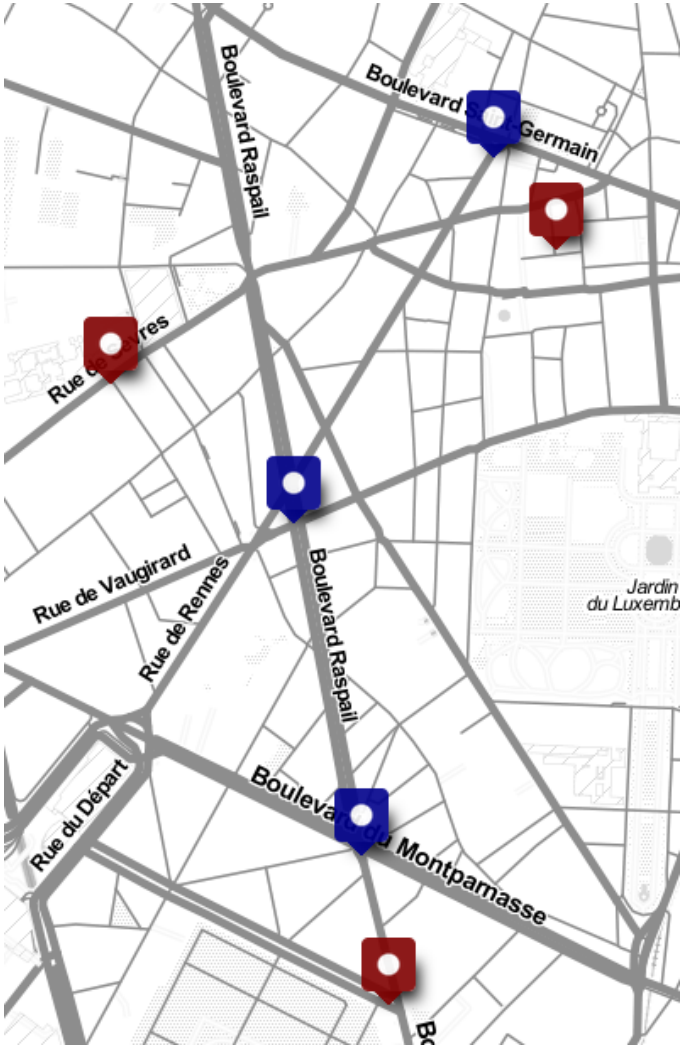


Protecting location traces

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

Distance between tuples:

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



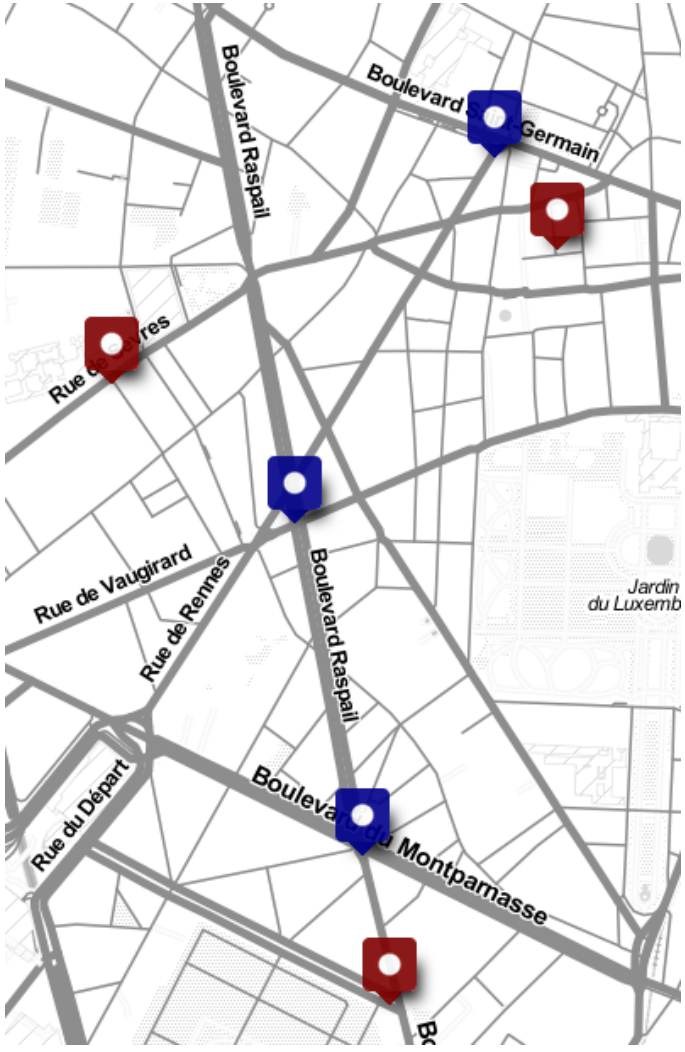
Independent Mechanism

apply noise to each point

n ϵ_N d_∞-private

works on any trace

budget is linear on n



Predictive Mechanism

prediction function

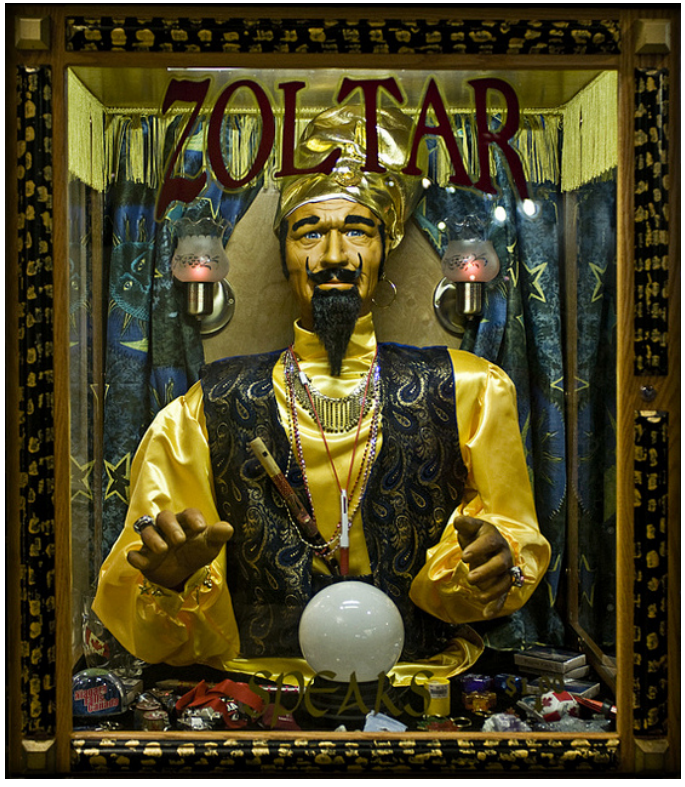
based on public info

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

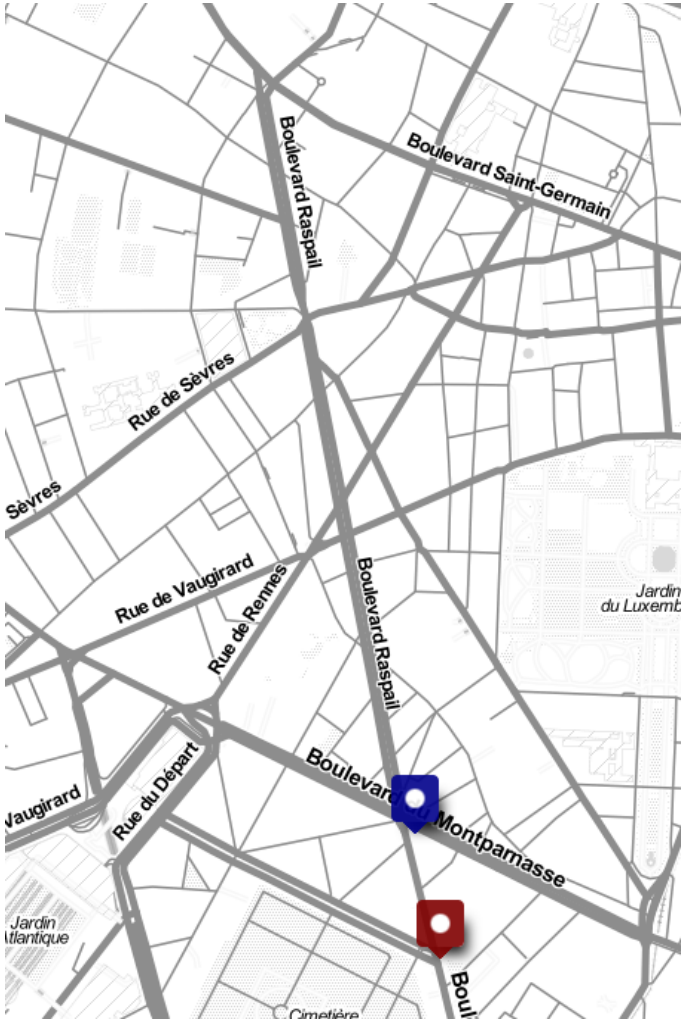
based on public info

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

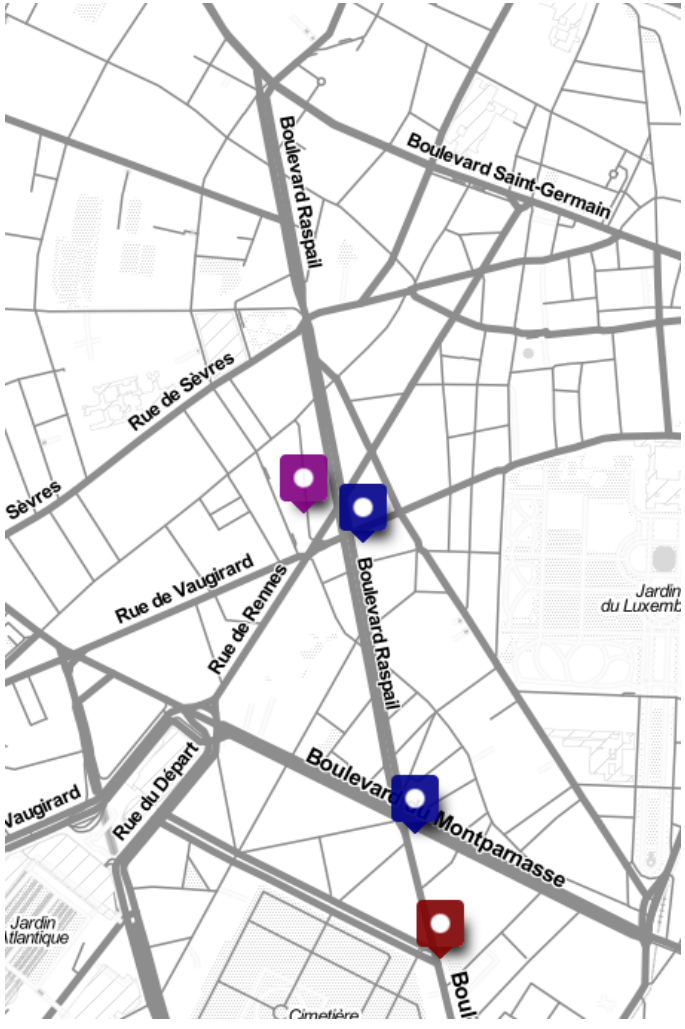
based on public info

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

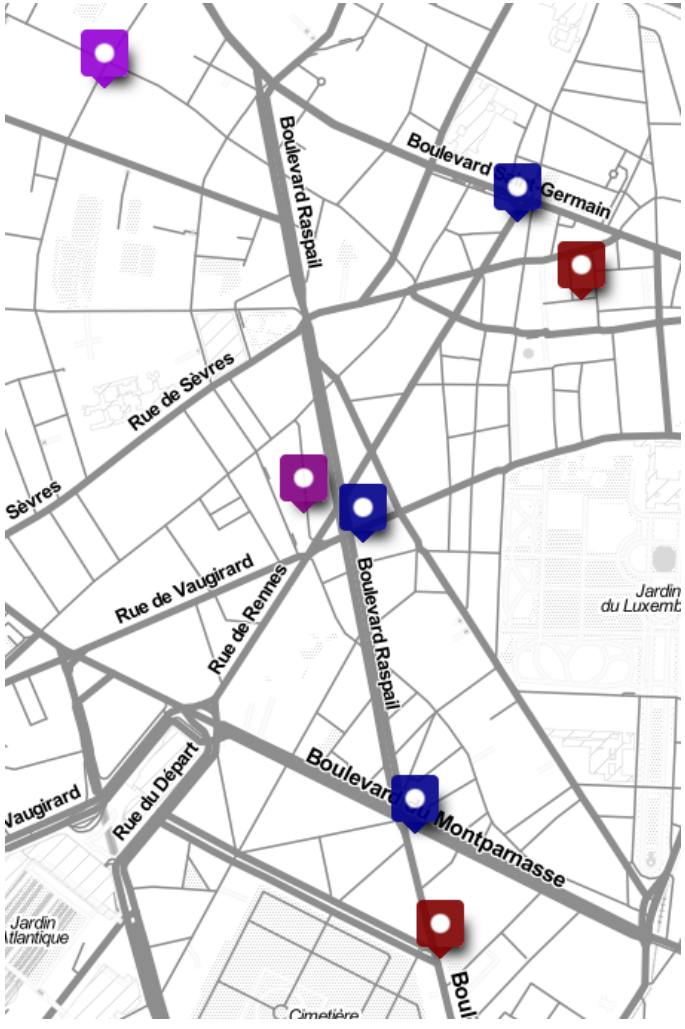
based on public info

obtain point $\tilde{z}_i$

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

yes: report $\tilde{z}_i$

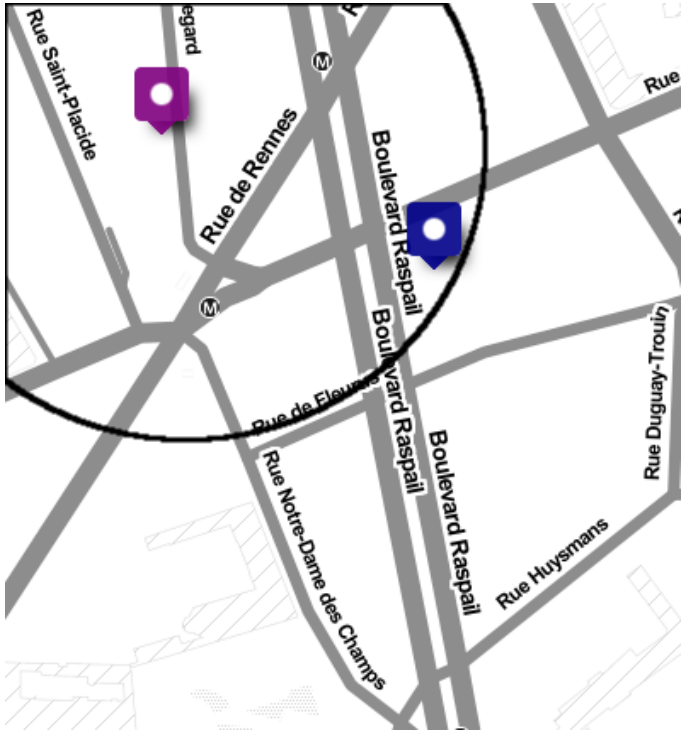
no: add new noise to x_i



Testing the prediction

Deterministic test

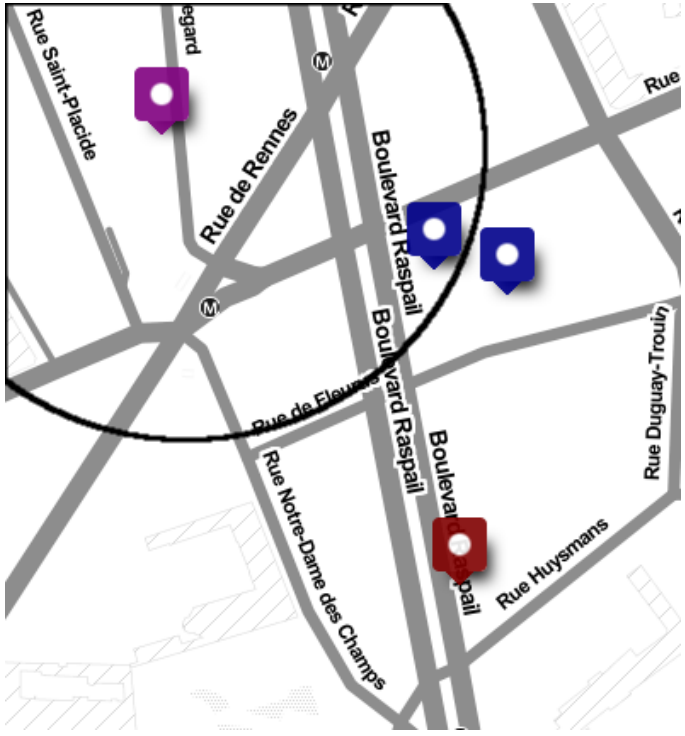
breaks privacy



Testing the prediction

Deterministic test

breaks privacy



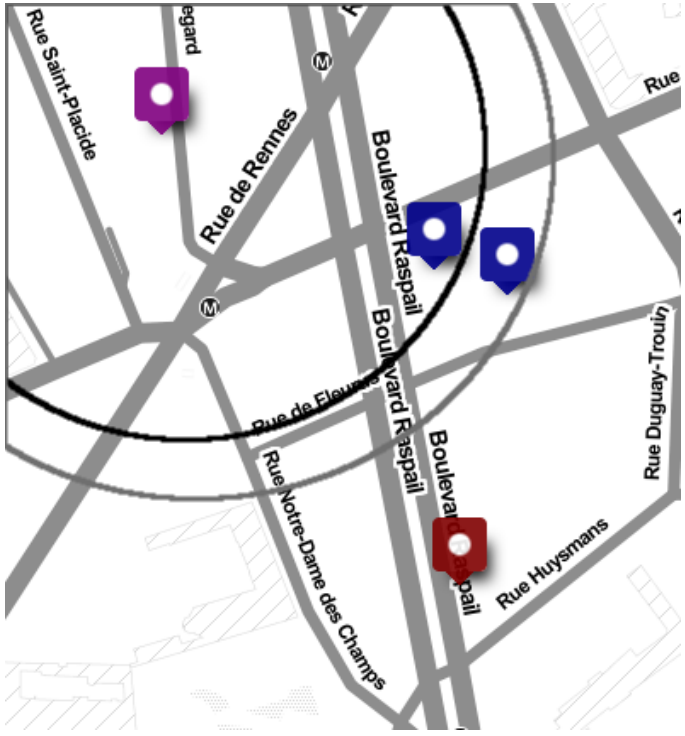
Testing the prediction

Deterministic test

breaks privacy

D-Private test

use a noisy border for the test



Testing the prediction

Deterministic test

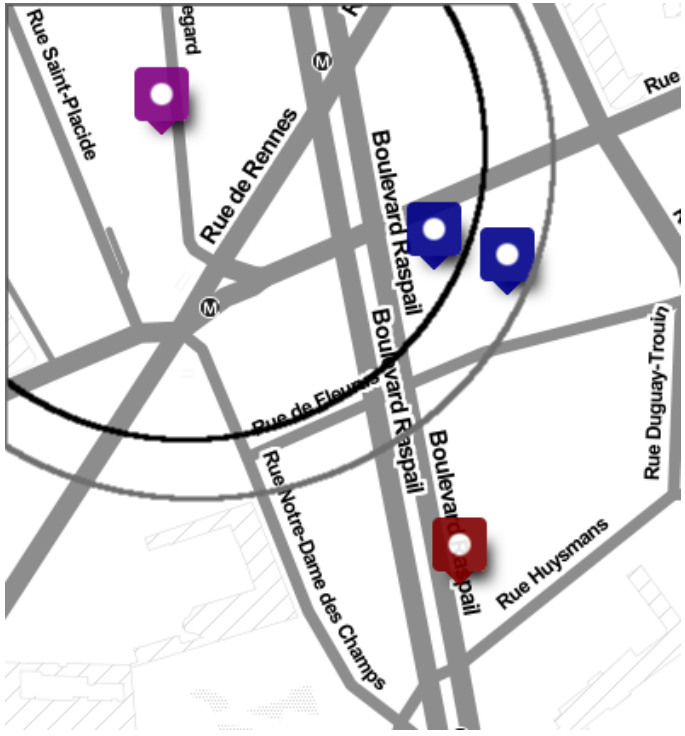
breaks privacy

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)



Budget Managers

Parameters

Local: $(\epsilon_\theta, \epsilon_N, l)$

Global: (ϵ, α, n)

Budget Manager: Global $\rightarrow$ Local

Privacy

ϵ is fixed. We propose two strategies:

Fixed Accuracy

What is saved is spent to increase n

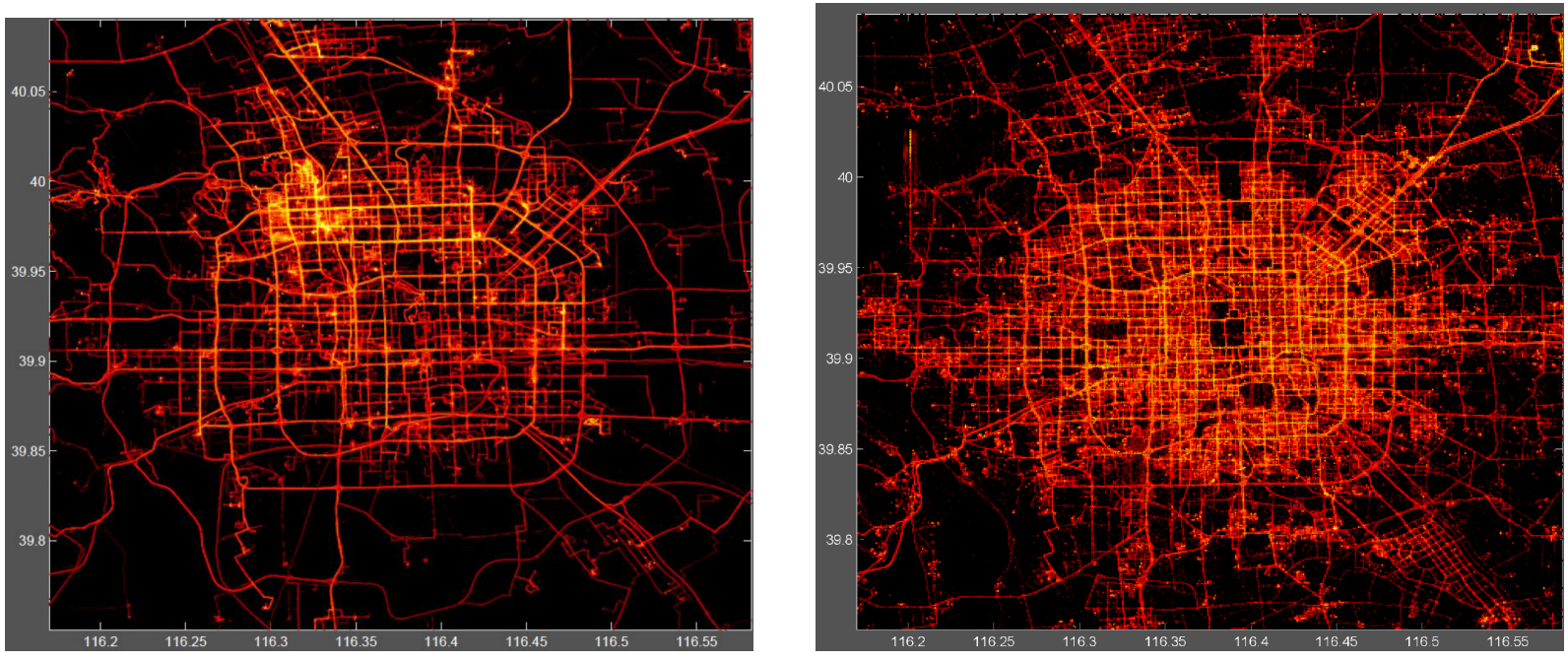
Fixed Rate

What is saved is spent to decrease α

Parrot prediction - simple yet effective

repeats the last observable

Evaluation using Geolife and TDrive



up to 65% improvement in budget consumption

up to 45% improvement in expected error

(In)Distinguishability Metric

What is it that you want to be similar to?

Euclidean Metric

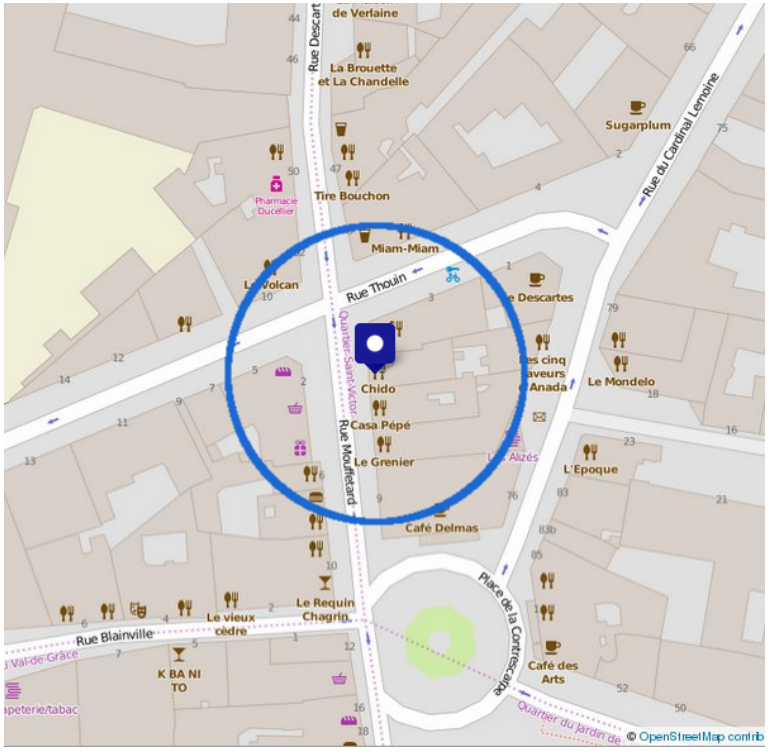
$$d_{\mathcal{X}}(x, x') = \epsilon \, d_E(x, x')$$

Space is privacy

ϵ tunes how much

Requirement

I want to be indistinguishable from a certain amount of space.



Euclidean Metric

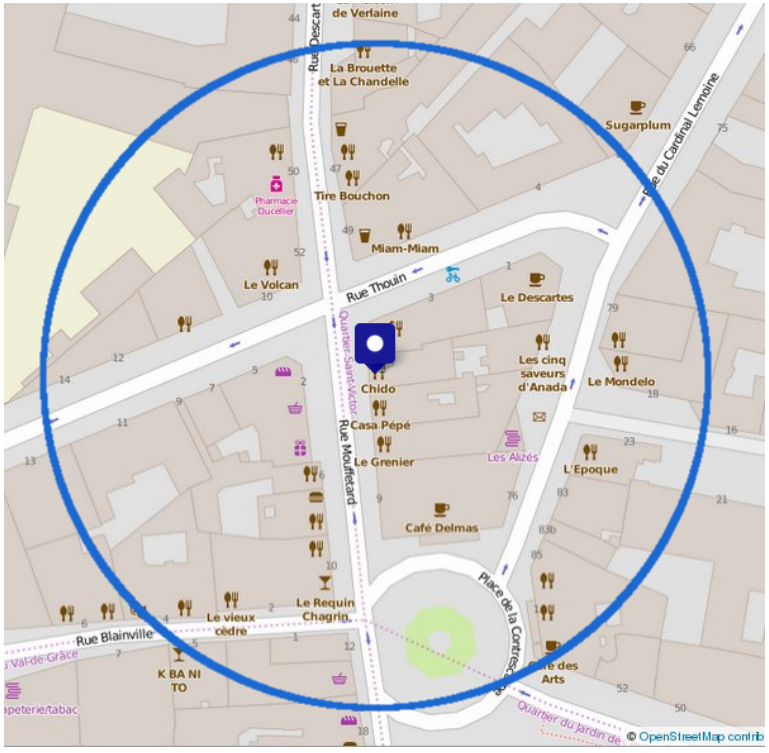
$$d_{\mathcal{X}}(x, x') = \epsilon \, d_E(x, x')$$

Space is privacy

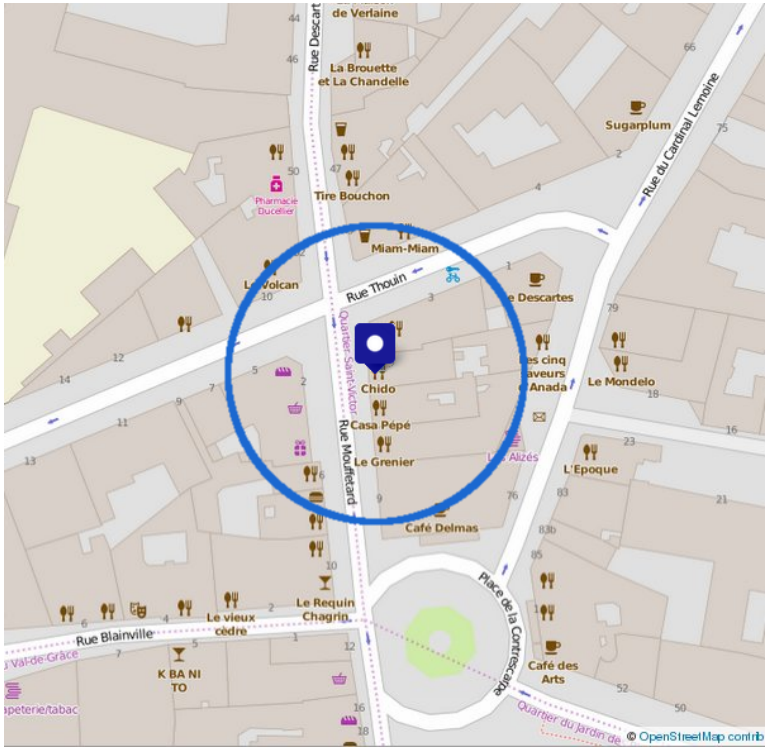
ϵ tunes how much

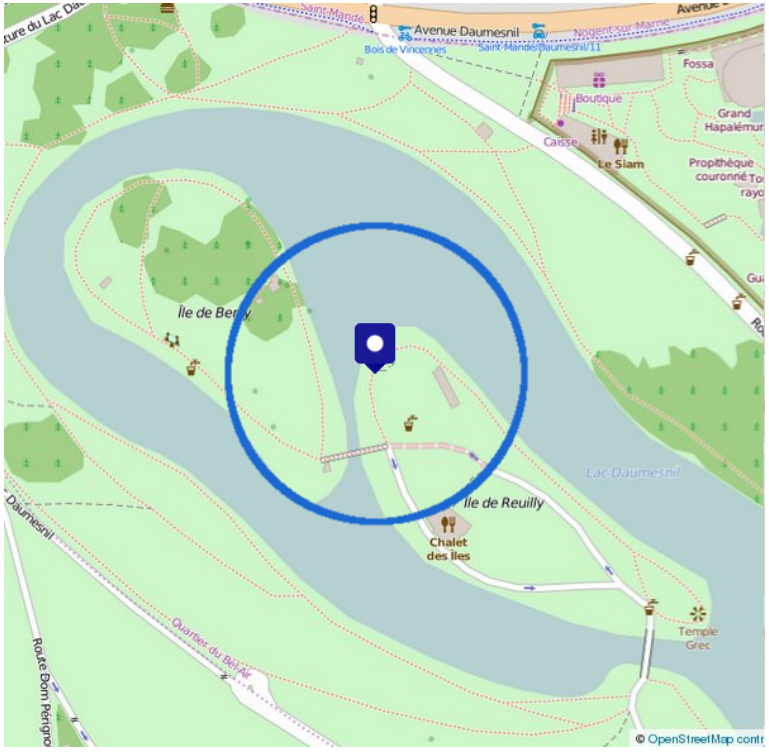
Requirement

I want to be indistinguishable from a certain amount of space.



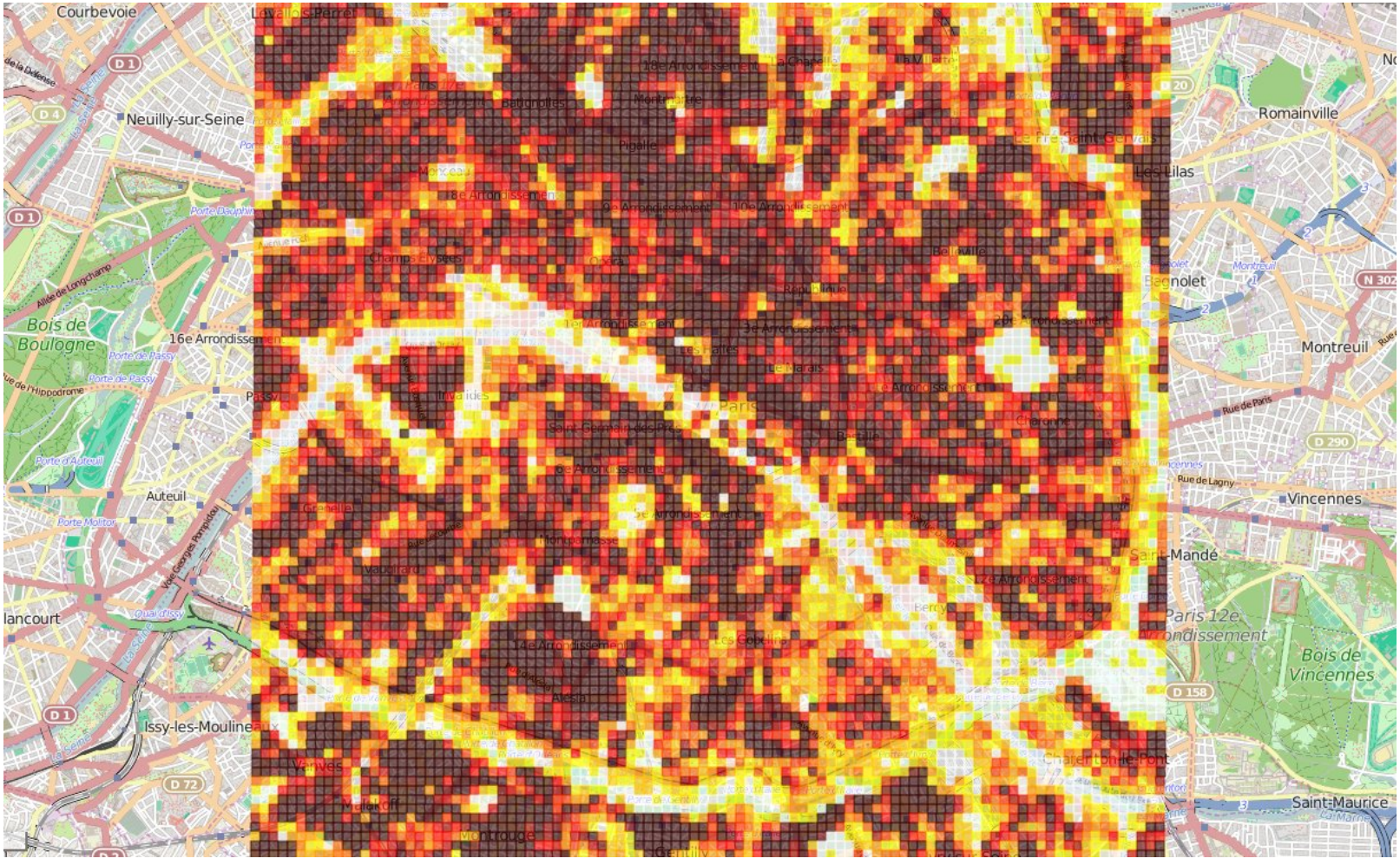
Not adaptable





Privacy Mass from OpenStreetMap

Privacy Mass from OpenStreetMap



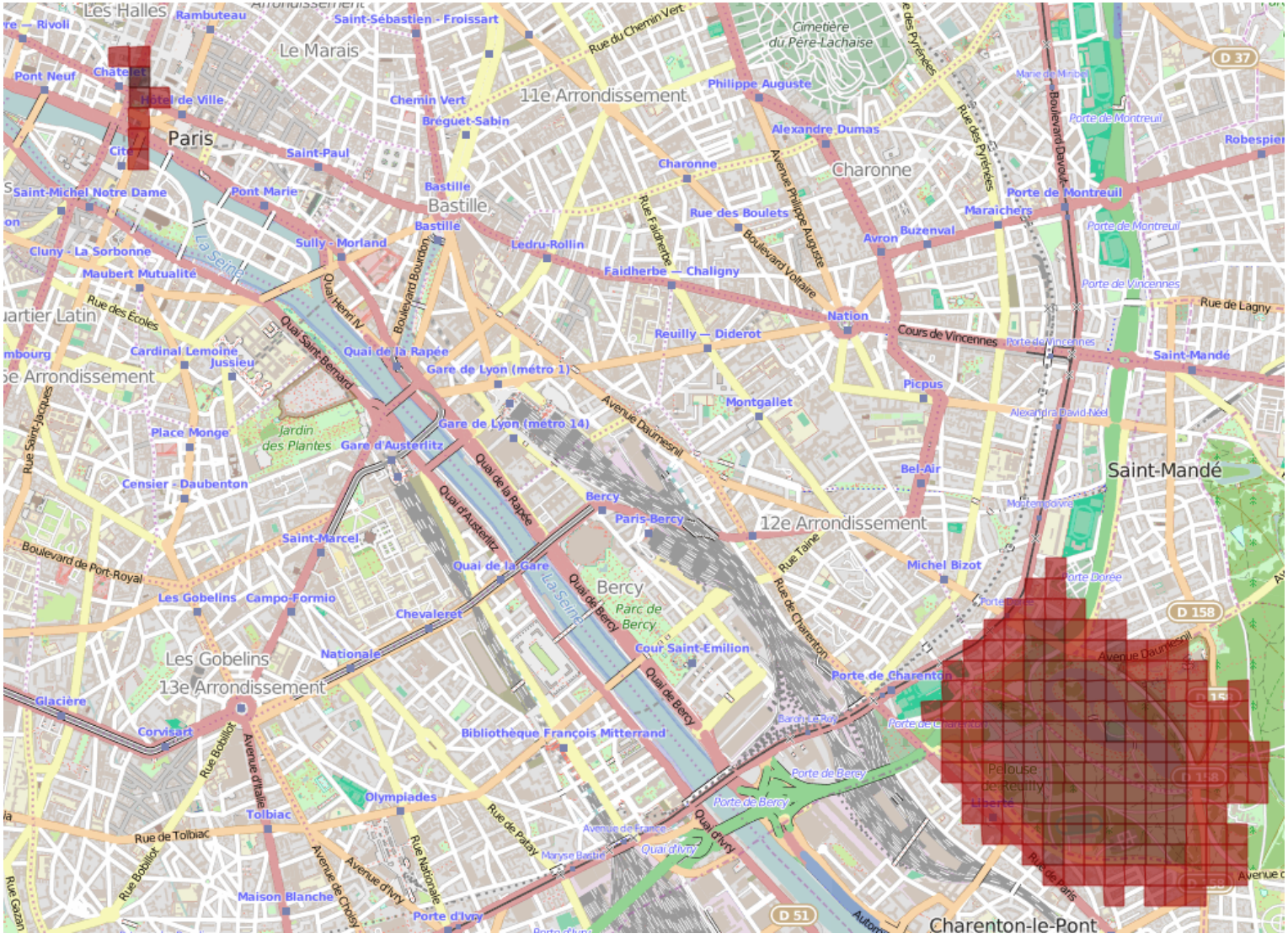
Elastic Mechanism

Requirement: I want to be indistinguishable from a "certain amount" of privacy mass

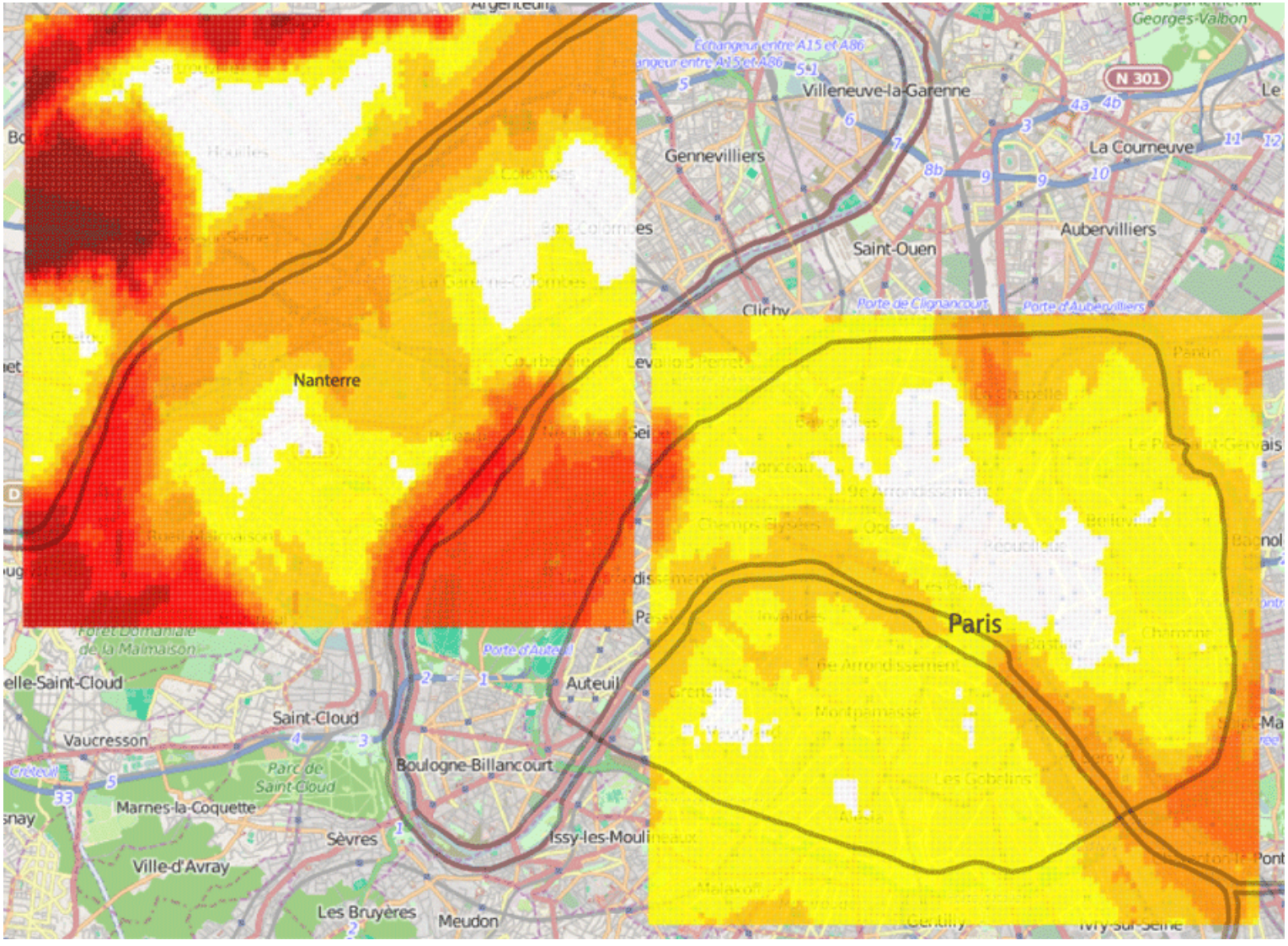
Scalable graph-based algorithm to build a metric from the requirement

Elastic Mechanism = Elastic Metric + Exponential Mechanism

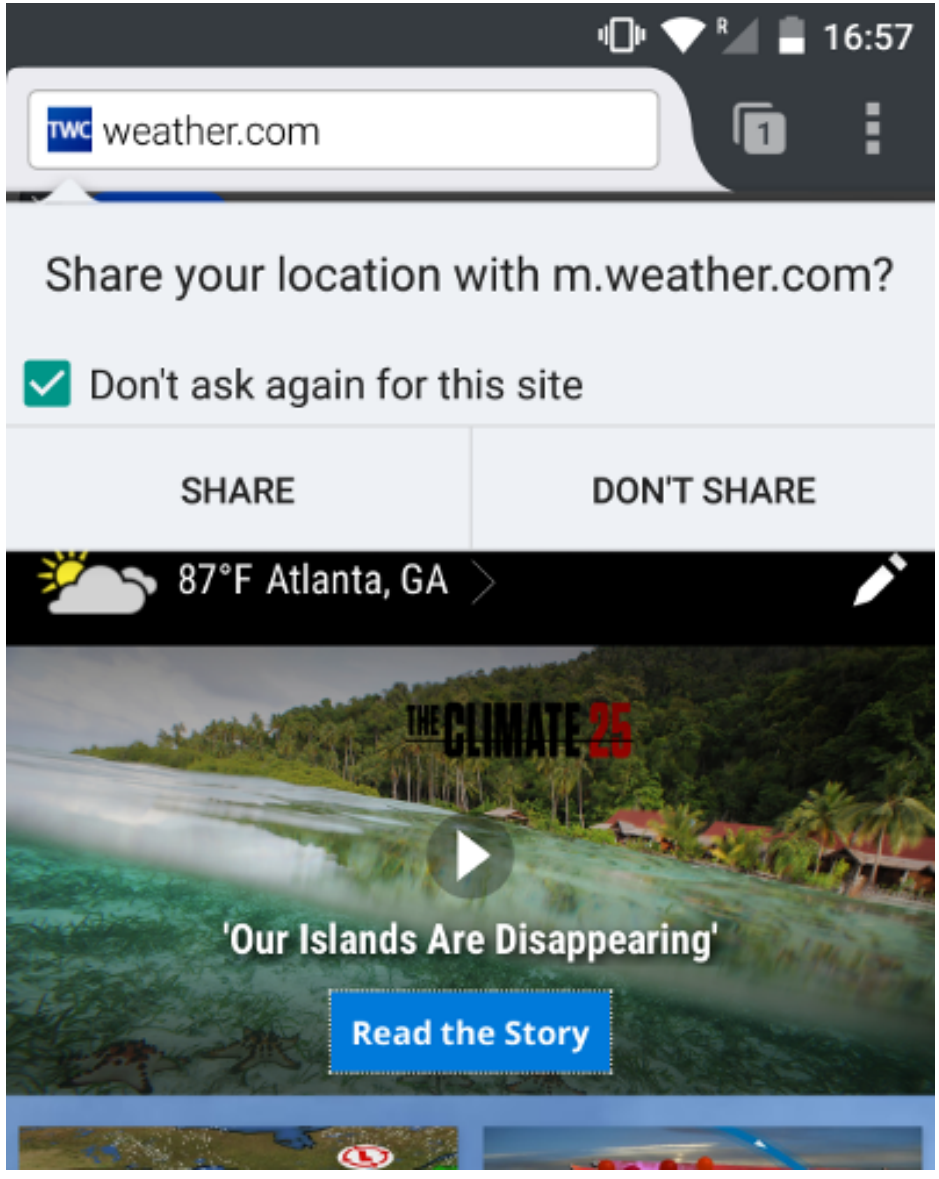
Elastic Mechanism



Elastic Mechanism



Location Guard



Location Guard: goals

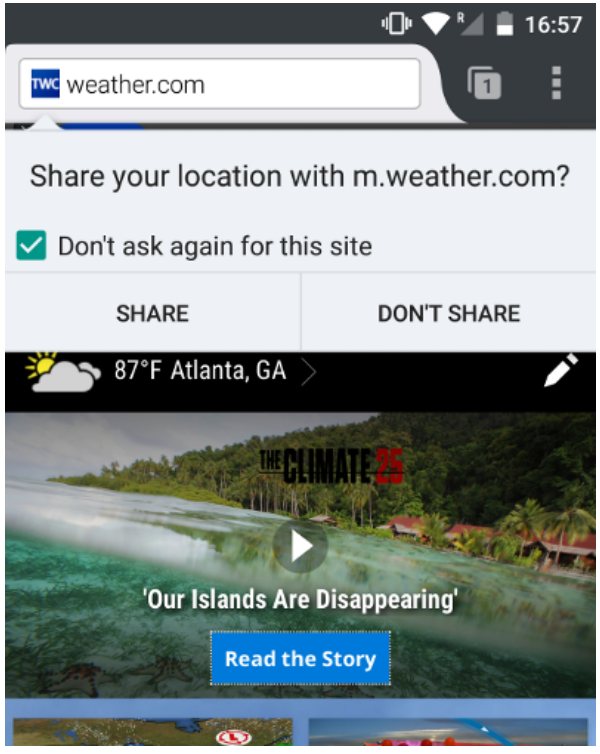
- Provide a simple practical solution
- enough so that common people actually use it
- Understandable, configurable by human beings
- Application-agnostic

    Browser level

    OS-level on smartphones (problem: rooting the phone)

W3C Geolocation API

```
navigator.geolocation.getCurrentPosition(function(pos)
    alert("Latitude: " + pos.coords.latitude +
          "Longitude:" + pos.coords.longitude);
);
```



Location Guard

- Intercept the javascript call

  Content-script, running in separate javascript enviroment

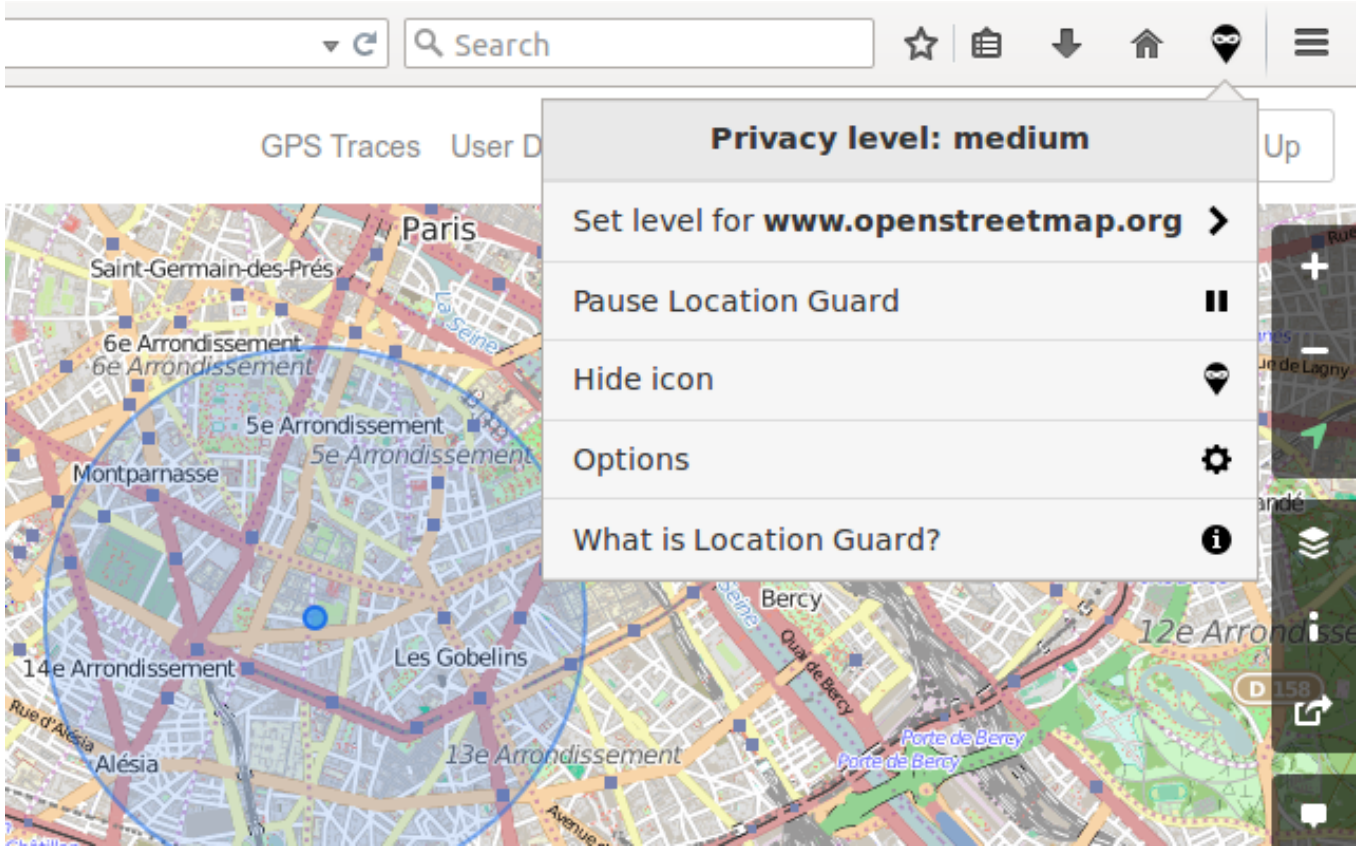
  Inject code in the page, replace `navigator.geolocation`

- Add noise, return the noisy location to the page

- Transparent to the user

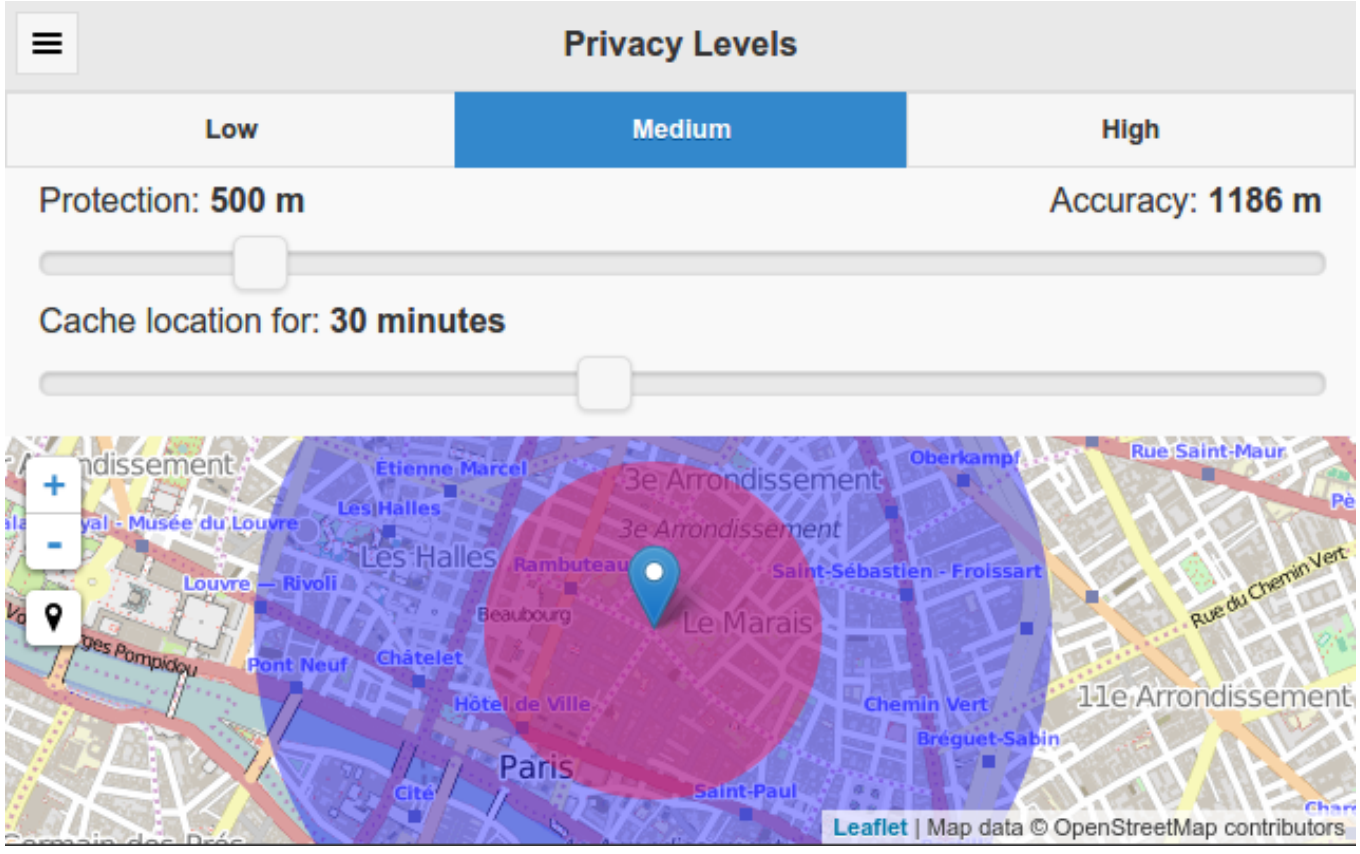
`github.com/chatziko/location-guard`

User interfaces are hard



No initial setup, user configuration if needed

User interfaces are hard



No initial setup, user configuration if needed

User adoption

Timeline

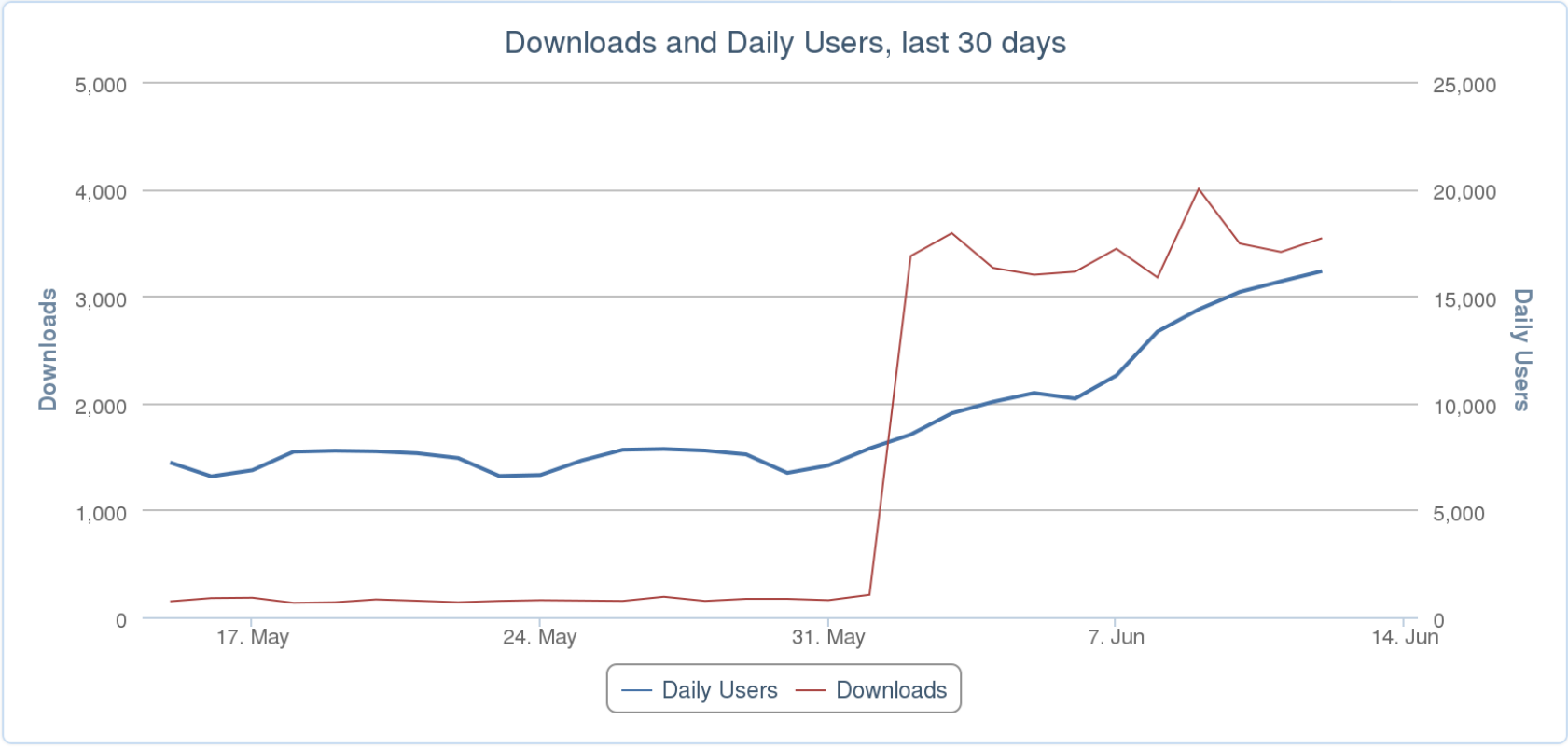
- Nov 2013: Chrome
- Jul 2014: Firefox
- Feb 2015: Firefox Mobile
- Feb 2015: Opera

Currently: 60k+ active users

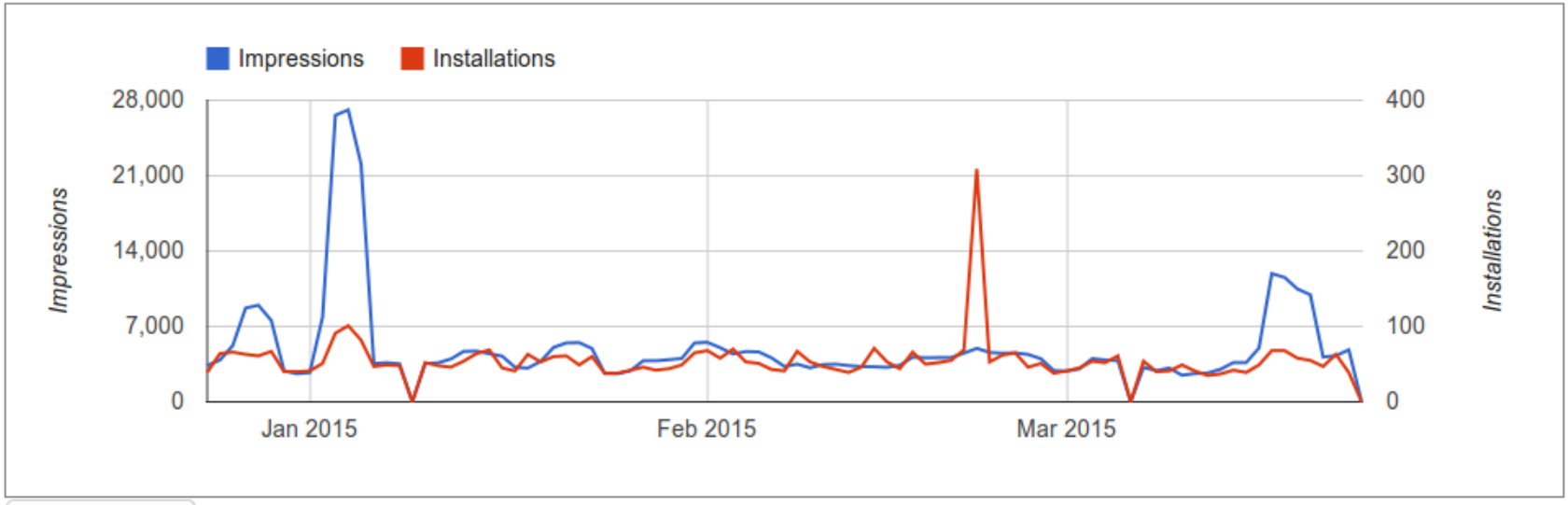
- Chrome: 16469 active
- Firefox: 38601 active
- Firefox Mobile: 2120 active
- Opera: 8282 downloads

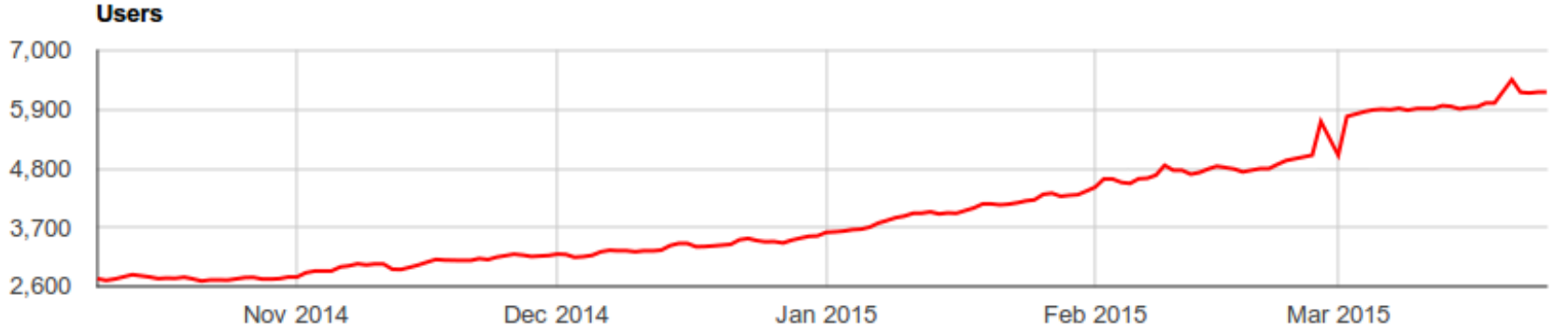
Pick of the month for June 2015

Pick of the month for June 2015



Chrome: linear growth





Reviews

5 stars

Works as advertised. [...] I like the idea of having a fixed location instead of just adding noise to the geo.

Reviews

1 star

After installation, I can not find any icon / starting point on FireFox!

We added a demo page.

Reviews

1 star

It doesn't work, weather.com and wunderground.com hit on my zip code immediately. I restarted the computer and it still doesn't work.

Reviews

5 stars

Very awesome idea, I didn't try this but I think it's a great idea.

Reviews

5 stars

I'm not a fan of using the NSA Ops1 building set as the default fixed location. I know it's a cheeky joke but ...

We changed it to the Manra island in the Pacific.

Reviews

4 stars

Can't do anything against google's geotargeting

Future directions

Try new mechanisms / techniques / . . .

Adapt the noise to the semantics of the location

Future directions

Try new mechanisms / techniques / ...

- Adapt the noise to the semantics of the location

Study the users' behaviour

- Collect data locally with the user's consent
- Run attacks/experiments/... locally
- User: visualize data, attacks, ...
- Research: collect the analysis' results

Future directions

Try new mechanisms / techniques / ...

- Adapt the noise to the semantics of the location

Study the users' behaviour

- Collect data locally with the user's consent
- Run attacks/experiments/... locally
- User: visualize data, attacks, ...
- Research: collect the analysis' results

More platforms (Firefox OS, Android, Microsoft Edge)

Future directions

Try new mechanisms / techniques / ...

- Adapt the noise to the semantics of the location

Study the users' behaviour

- Collect data locally with the user's consent
- Run attacks/experiments/... locally
- User: visualize data, attacks, ...
- Research: collect the analysis' results

More platforms (Firefox OS, Android, Microsoft Edge)

Questions?