Meeting on Formal Methods on Hiding and Disclosing Information
December 3rd, 2013

Protecting Location Privacy with an Optimal and Differentially Private Obfuscation Mechanism

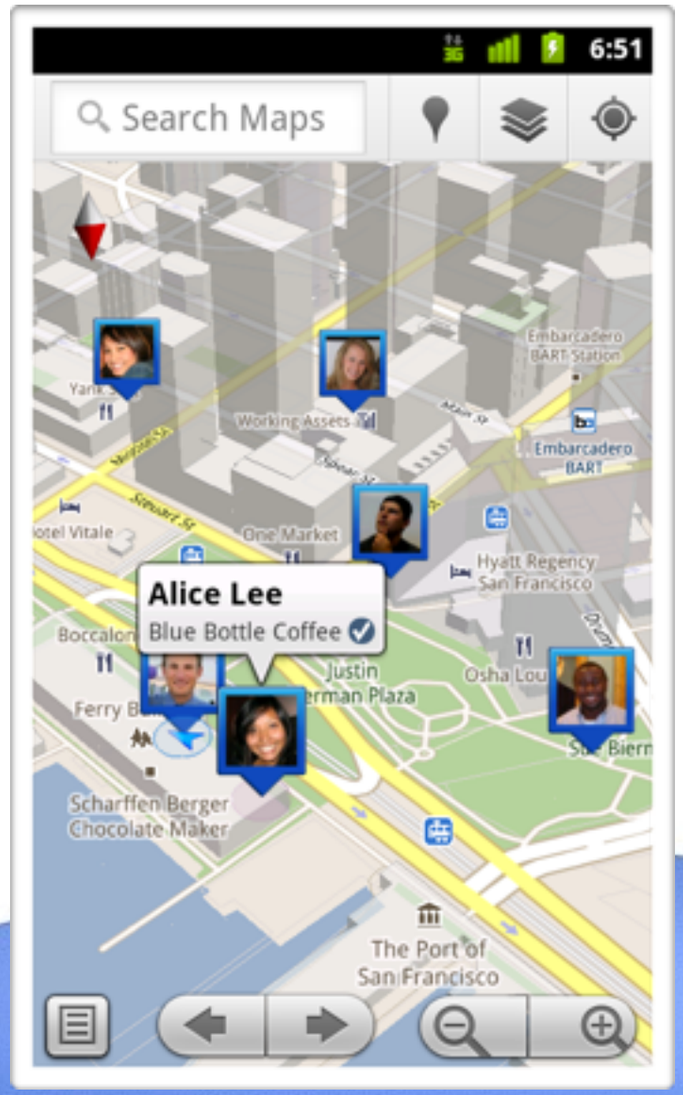
Nicolás E. Bordenabe

Joint work with:
Miguel Andrés, Kostas Chatzikokolakis, Catuscia Palamidessi

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.



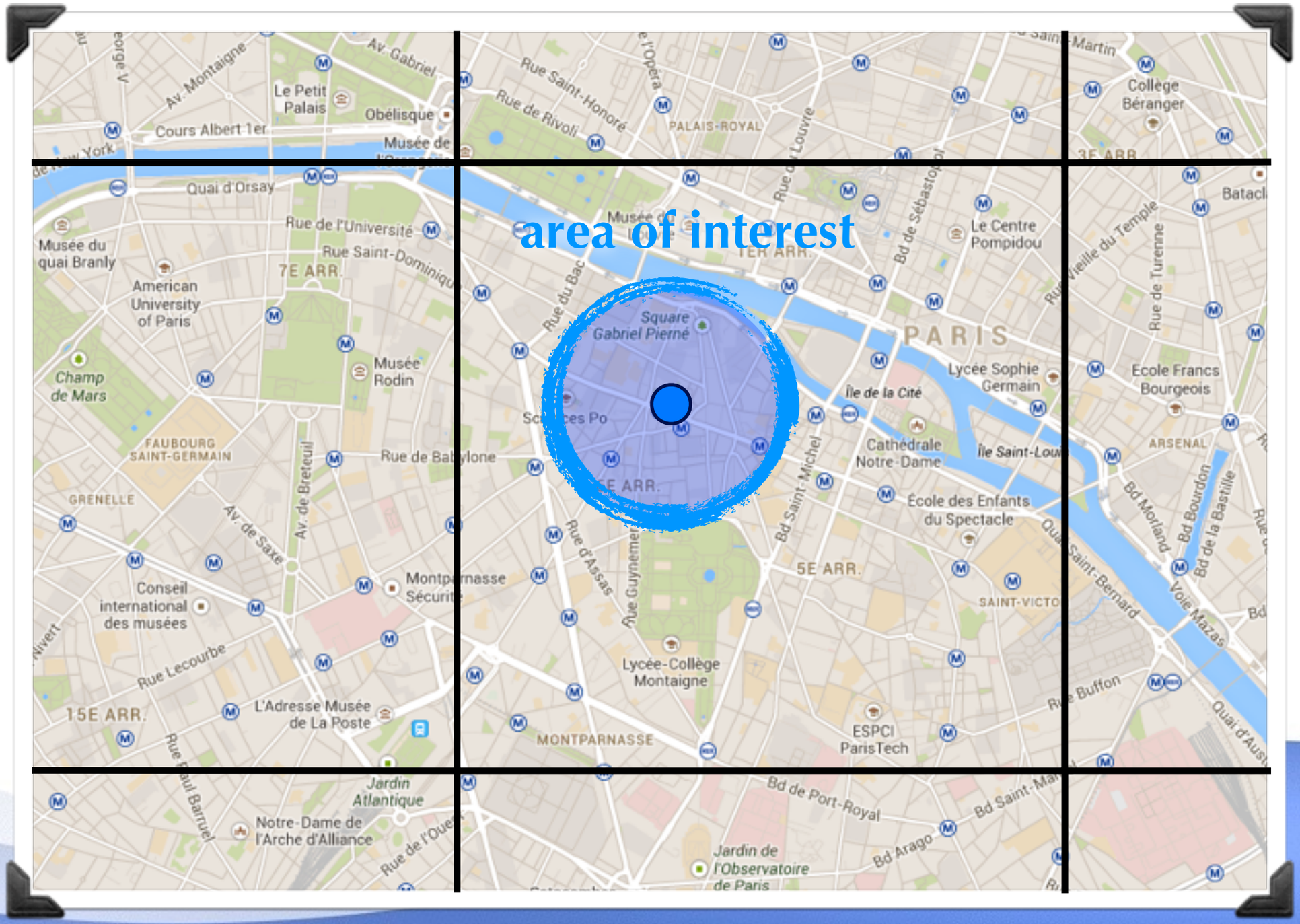
Location-Based Systems

- 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**.

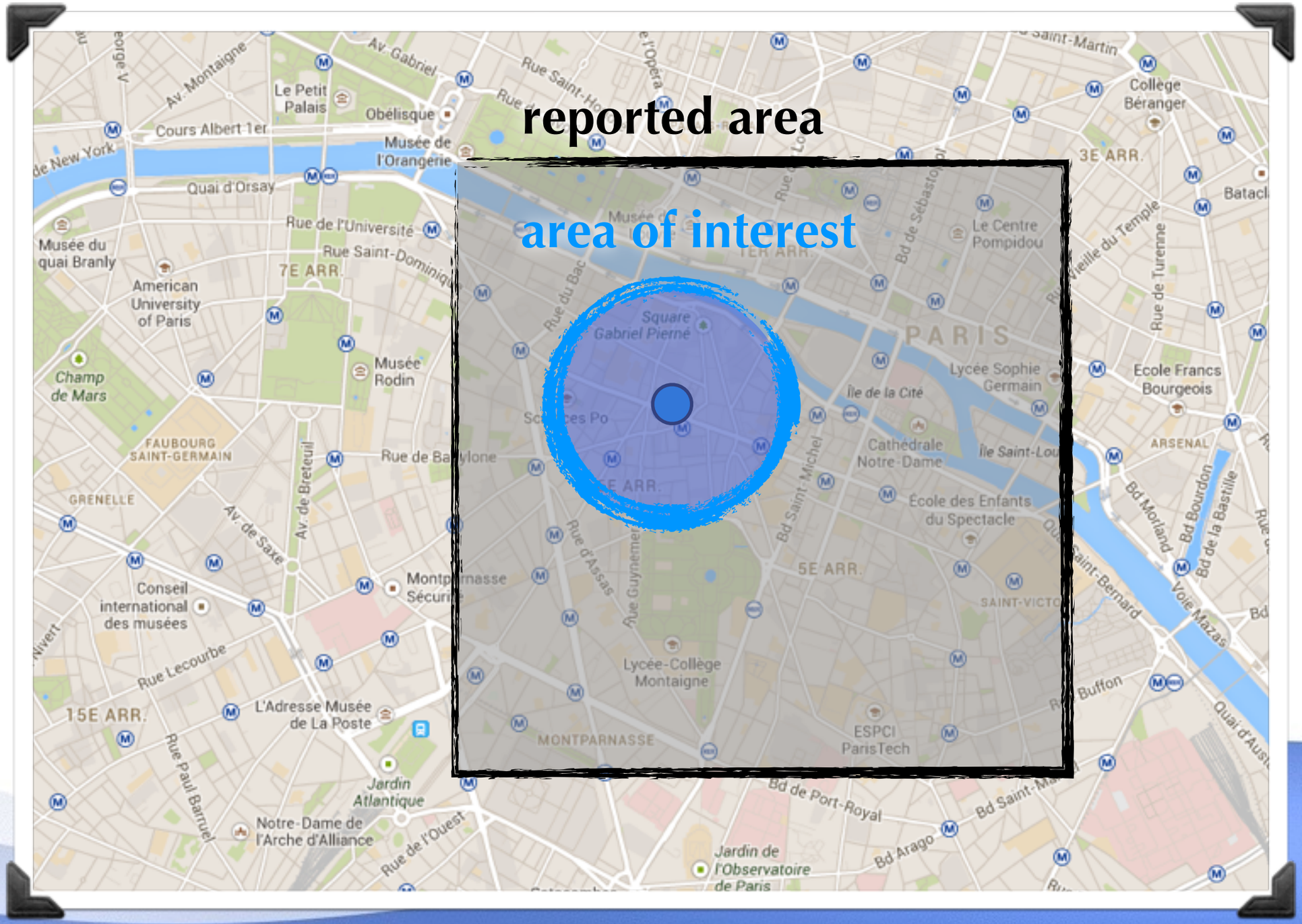
Example

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

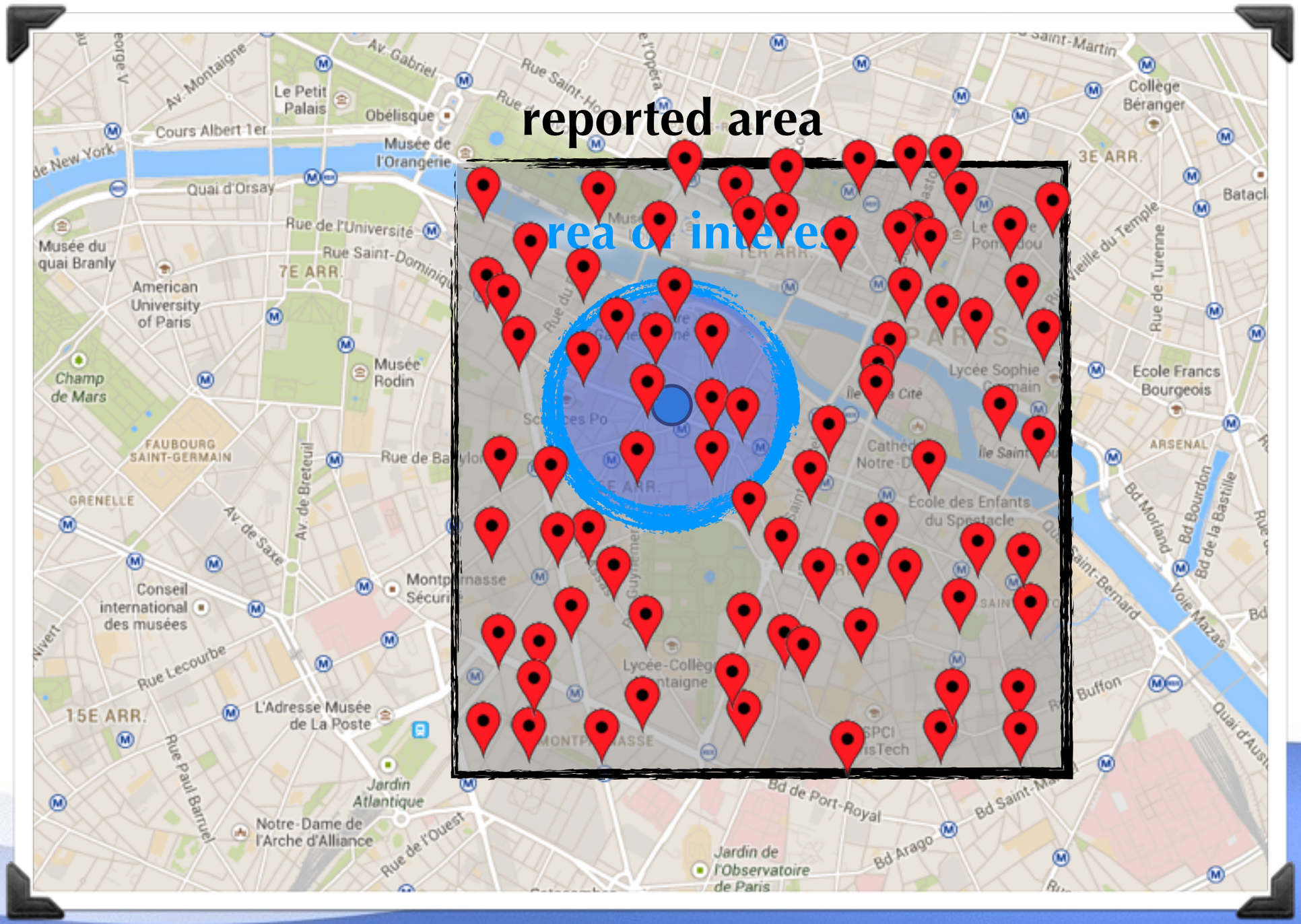
Cloaking


area of interest
Av. Montaigne
Av. Gabriel
Le Petit Palais
Obélisque
Cours Albert 1er
Musée de
Rue Saint-Honoré
Rue de Rivoli
PALAIS-ROYAL
Collège Béranger
Quai d'Orsay
Musée du quai Branly
Rue de l'Université
Rue Saint-Dominique
7E ARR.
American University of Paris
Champ de Mars
Musée Rodin
FAUBOURG SAINT-GERMAIN
Rue de Babylone
GRENELLE
Av. de Saxe
Av. de Breteuil
Conseil international des musées
Montparnasse Sécurité
Rue Lecourbe
15E ARR.
L'Adresse Musée de La Poste
Jardin Atlantique
Notre-Dame de l'Arche d'Alliance
Rue de l'Ouest
Rue Paul Barruel
Rue du Bac
Square Gabriel Pierné
Sciences Po
Rue d'Assas
Rue Guynemer
Lycée-Collège Montaigne
MONTPARNASSE
Bd de Port-Royal
Jardin de l'Observatoire de Paris
Bd Arago
Bd Saint-Michel
Rue du Louvre
Bd de Sébastopol
1ER ARR.
Le Centre Pompidou
PARIS
Île de la Cité
Cathédrale Notre-Dame
Lycée Sophie Germain
École des Enfants du Spectacle
5E ARR.
SAINT-VICTO
ESPCI ParisTech
Rue Vieille du Temple
Rue de Turenne
Ecole Francs Bourgeois
ARSENAL
Bd Morland
Bd Bourdon
Bd de la Bastille
Voie Mazas
Quai d'Austerlitz
Saint-Bernard
Rue Buffon

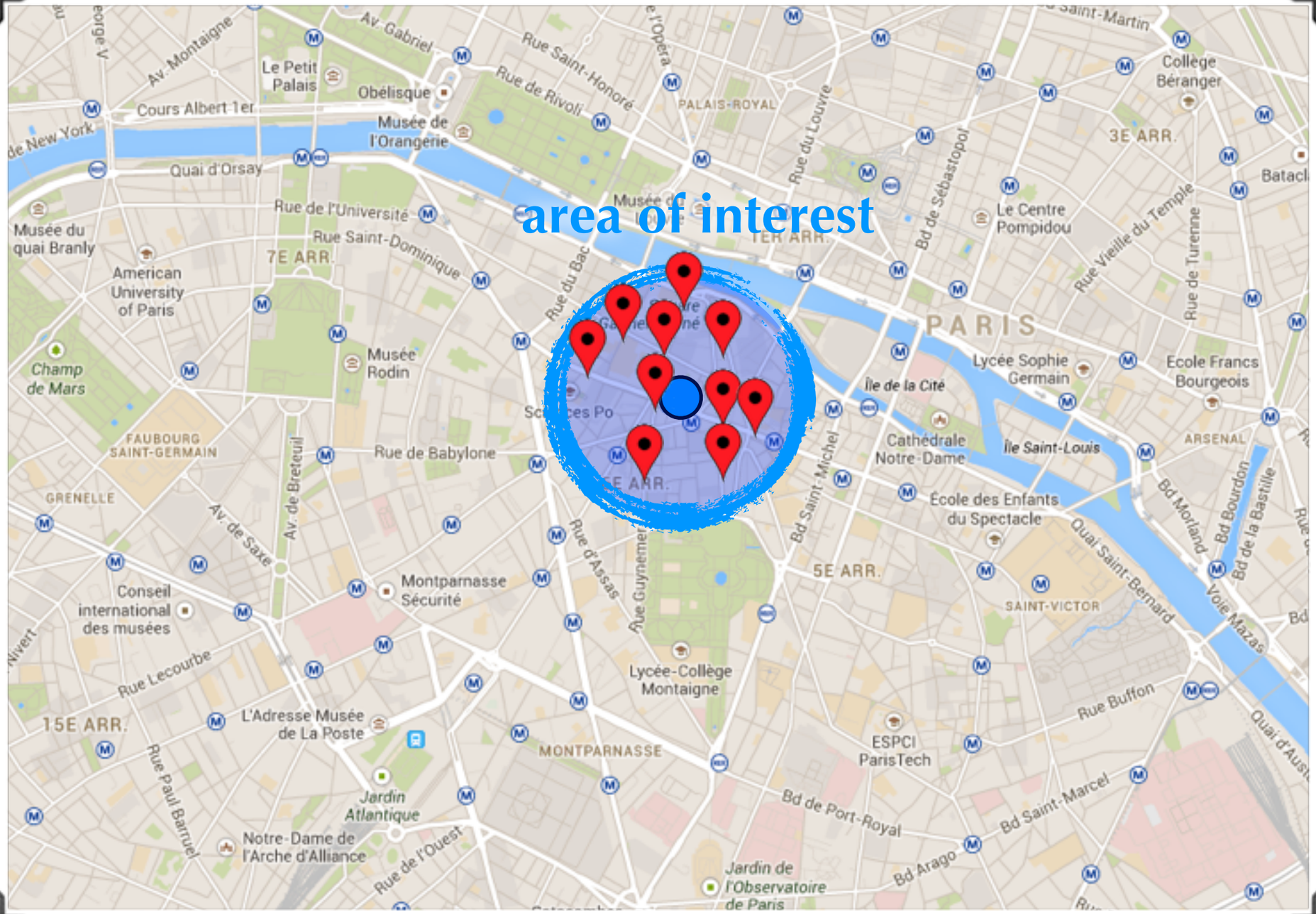
Cloaking


reported area
area of interest
Av. Gabriel
Av. Montaigne
Le Petit Palais
Obélisque
Cours Albert 1er
Musée de l'Orangerie
Quai d'Orsay
Rue de l'Université
Rue Saint-Dominique
Musée du quai Branly
American University of Paris
7E ARR.
Champ de Mars
Musée Rodin
FAUBOURG SAINT-GERMAIN
Rue de Babylone
Av. de Breteuil
GRENELLE
Av. de Saxe
Conseil international des musées
Rue Lecourbe
15E ARR.
L'Adresse Musée de La Poste
Rue Paul Barruel
Jardin Atlantique
Notre-Dame de l'Arche d'Alliance
Rue de l'Ouest
Square Gabriel Pierné
Sciences Po
Rue du Bac
Rue d'Assas
Rue Guynemer
Bd Saint-Michel
Lycée-Collège Montaigne
MONTPARNASSE
ESPCI ParisTech
Bd de Port-Royal
Jardin de l'Observatoire de Paris
Bd Arago
Bd de Sébastopol
Le Centre Pompidou
1ER ARR.
PARIS
Île de la Cité
Cathédrale Notre-Dame
Lycée Sophie Germain
École des Enfants du Spectacle
5E ARR.
Rue Saint-Martin
Collège Béranger
3E ARR.
Rue de Turenne
Boulevard Beaumarchais
Ecole Francs Bourgeois
ARSENAL
Bd Morland
Bd Bourdon
Bd de la Bastille
Voie Mazas
Quai d'Austerlitz
Buffon

Cloaking


reported area
area of interest
Av. Montaigne
Le Petit Palais
Obélisque
Av. Gabriel
Rue Saint-Honoré
Cours Albert 1er
Musée de l'Orangerie
Quai d'Orsay
Rue de l'Université
Rue Saint-Dominique
Musée du quai Branly
7E ARR.
American University of Paris
Champ de Mars
Musée Rodin
FAUBOURG SAINT-GERMAIN
Rue de Babylone
GRENELLE
Av. de Breteuil
Av. de Saxe
Conseil international des musées
Montparnasse
Rue Lecourbe
15E ARR.
L'Adresse Musée de La Poste
Rue Paul Barruel
Jardin Atlantique
Notre-Dame de l'Arche d'Alliance
Rue de l'Ouest
Lycée-Collège
MONTPARNASSE
Bd de Port-Royal
Bd Arago
Jardin de l'Observatoire de Paris
Lycée Sophie Germain
Île de la Cité
Cathédrale Notre-Dame
Île Saint-Louis
École des Enfants du Spectacle
3E ARR.
Collège Béranger
Rue de Turenne
Rue Vieille du Temple
ARSENAL
Bd Morland
Bd Bourdon
Bd de la Bastille
Voie Mazas
Quai d'Austerlitz
Buffon
Ecole Francs Bourgeois

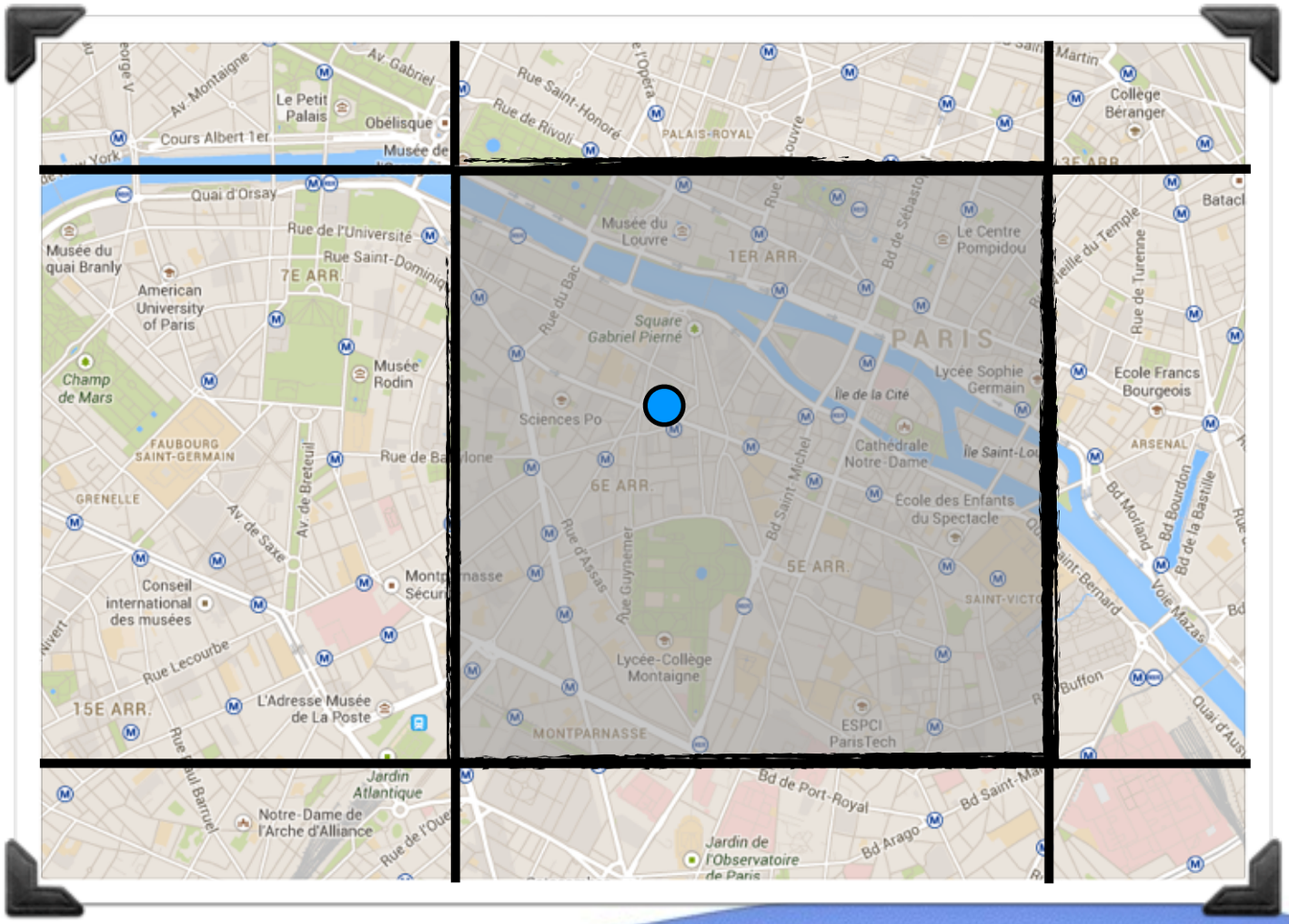
Cloaking


area of interest
Av. Gabriel
Rue Saint-Honoré
Rue de Rivoli
Av. Montaigne
Le Petit Palais
Obélisque
Cours Albert 1er
Musée de l'Orangerie
Quai d'Orsay
Rue de l'Université
Rue Saint-Dominique
Musée du quai Branly
American University of Paris
7E ARR.
Champ de Mars
Musée Rodin
FAUBOURG SAINT-GERMAIN
Rue de Babylone
GRENELLE
Av. de Breteuil
Av. de Saxe
Conseil international des musées
Montparnasse Sécurité
Rue Lecourbe
15E ARR.
L'Adresse Musée de La Poste
Jardin Atlantique
Notre-Dame de l'Arche d'Alliance
Rue de l'Ouest
Rue Paul Barruel
MONTPARNASSE
Rue d'Assas
Rue Guynemer
Lycée-Collège Montaigne
Jardin de l'Observatoire de Paris
Bd de Port-Royal
Bd Arago
ESPCI ParisTech
Rue du Bac
Rue du Louvre
PALAIS-ROYAL
1ER ARR.
Bd de Sébastopol
Le Centre Pompidou
Rue Vieille du Temple
Rue de Turenne
3E ARR.
Collège Béranger
Batacl
PARIS
Lycée Sophie Germain
Île de la Cité
Cathédrale Notre-Dame
Île Saint-Louis
Ecole Francs Bourgeois
ARSENAL
Bd Saint-Michel
École des Enfants du Spectacle
5E ARR.
Quai Saint-Bernard
Bd Morland
Bd Bourdon
Bd de la Bastille
Voie Mazas
SAINT-VICTOR
Rue Buffon
Quai d'Austerlitz
Bd Saint-Marcel

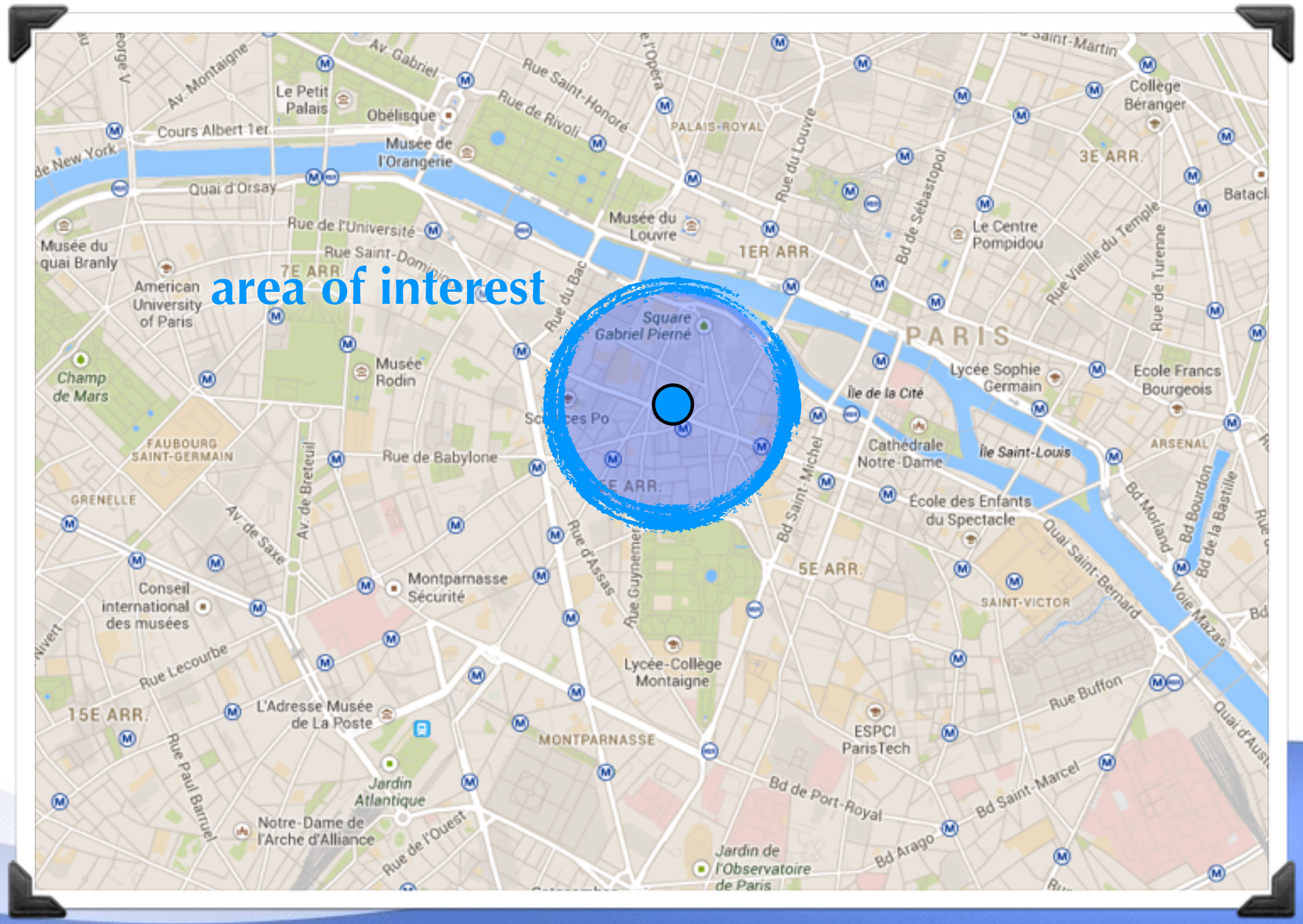
Cloaking

Drawback:

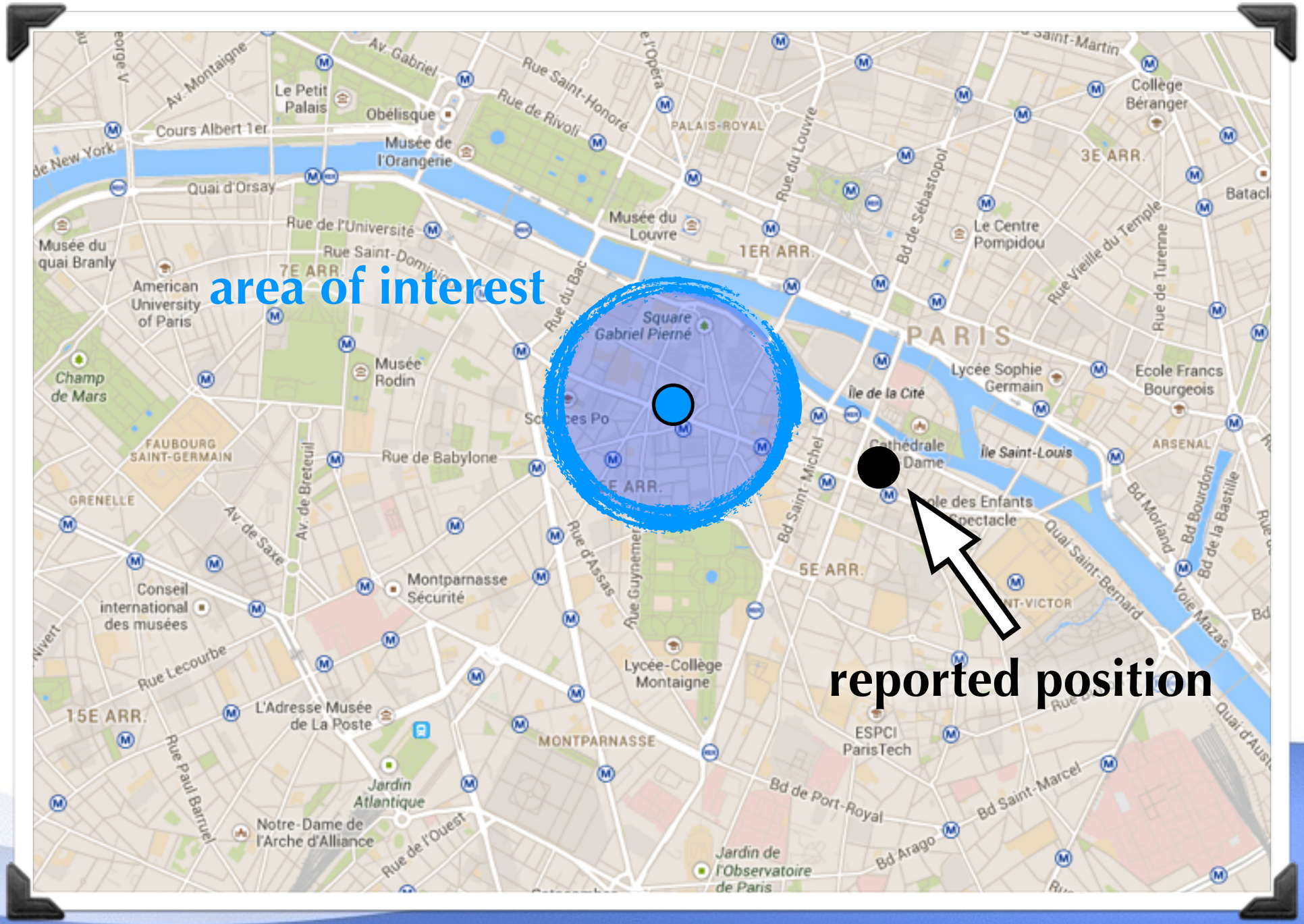
- Movement between **borders**.
- Sensitive to **prior information**.



Obfuscation

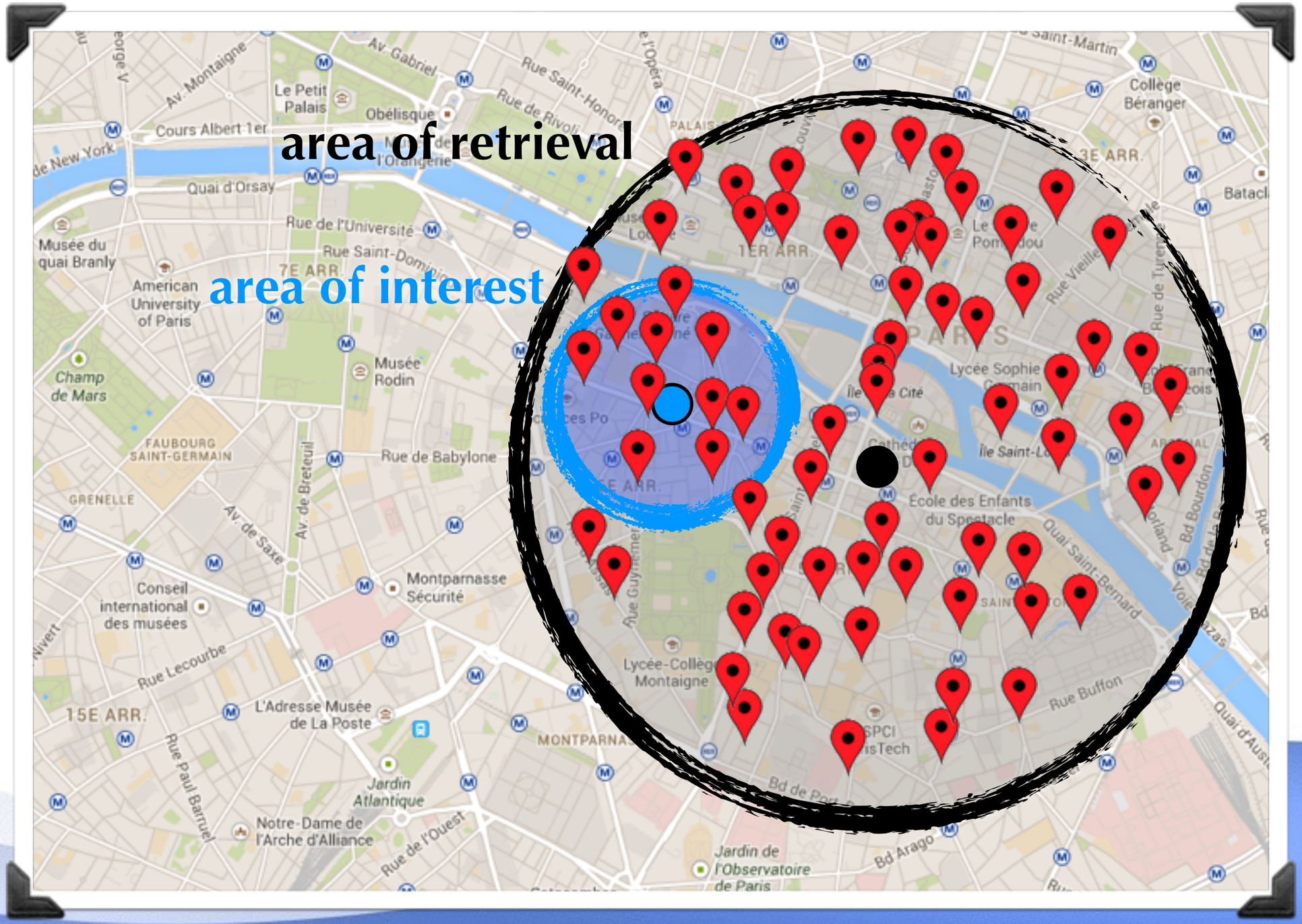


Obfuscation

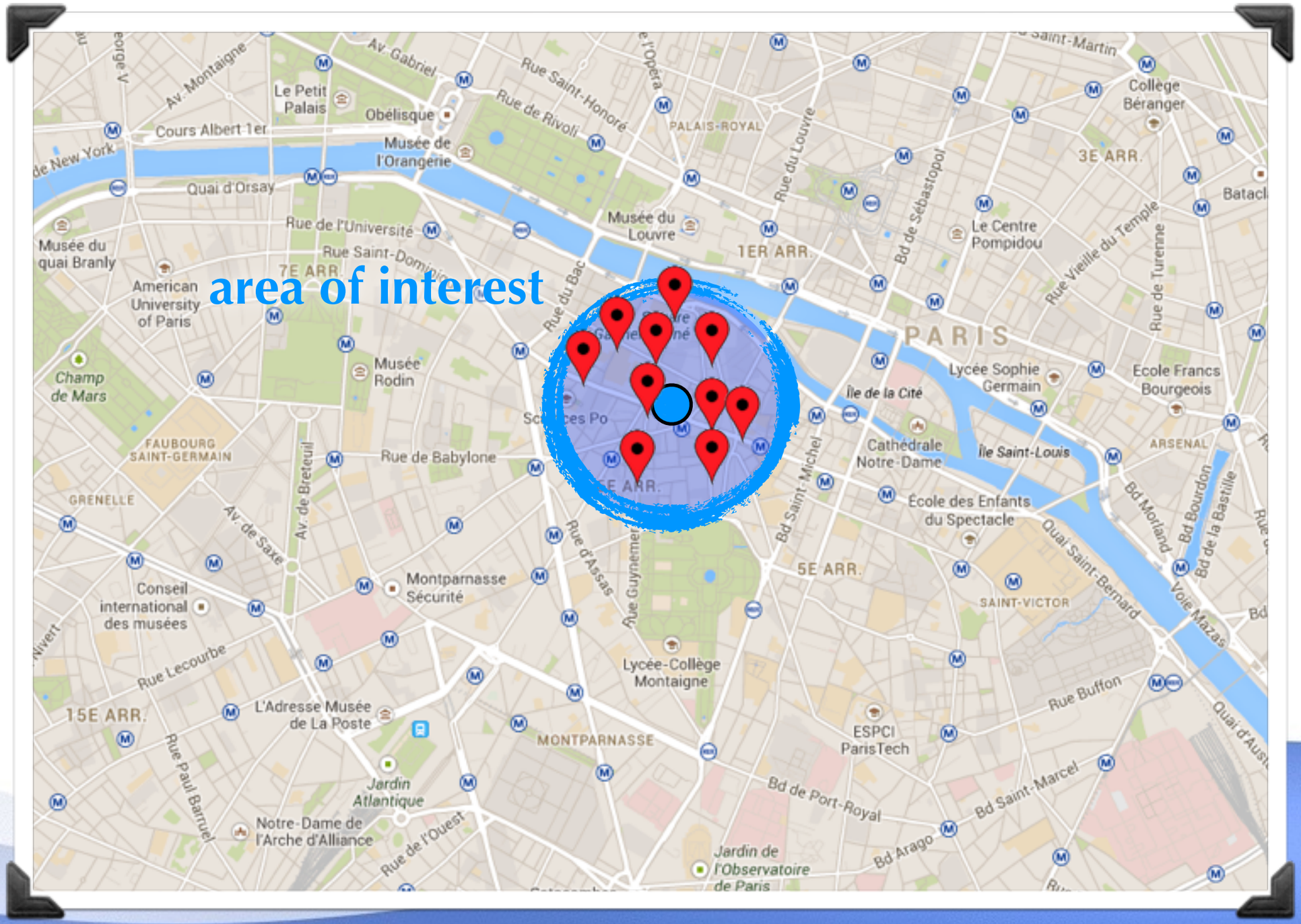


Obfuscation

Obfuscation



Obfuscation



Obfuscation

Shokri et al., CCS 2012:

- Mechanism with **optimal privacy** for a **fixed quality of service**.
- Privacy measure: **expected error of the adversary** under a given user profile.
- Quality measure: **expected distance** between the reported location and the real one.
- Mechanism obtained by solving an optimization problem:
  - Maximize privacy.
  - Constraint on the quality loss.

Obfuscation

Shokri et al., CCS 2012:

- Mechanism with **optimal privacy** for a **fixed quality of service**.
- Privacy measure: **expected error of the adversary** under a given user profile.
- Quality measure: **expected dista** between the reported location and the real one.
- Mechanism obtained by sol

  - Maximize privacy.
  - Constraint on the quality loss.

depends on a prior distribution over the locations

The Goals

1. We want an **obfuscation mechanism**.

2. Formal privacy definition, **independent from prior information**.

3. **Optimal** utility for a given user, hard constraint on the privacy.

Differential Privacy

An attacker should not be able to learn personal information from statistical queries.

Name	Age
Eddard	35
Jon	15
Arya	9
Jaime	32
Joffrey	12
Tyrion	24
Daenerys	14

How many users are 16 or older ?

Is Jon 16 or older ?

Differential Privacy

An attacker should not be able to learn personal information from statistical queries.

Name	Age
Eddard	35
Jon	15
Arya	9
Jaime	32
Joffrey	12
Tyrion	24
Daenerys	14

How many users are 16 or older ?

→ 3

Is Jon 16 or older ?

No

Differential Privacy

An attacker should not be able to learn personal information from statistical queries.

Name	Age
Eddard	35
Jon	15
Arya	9
Jaime	32
Joffrey	12
Tyrion	24
Daenerys	14

How many users are 16 or older ?

➡ 3 ➡ 3+**x**

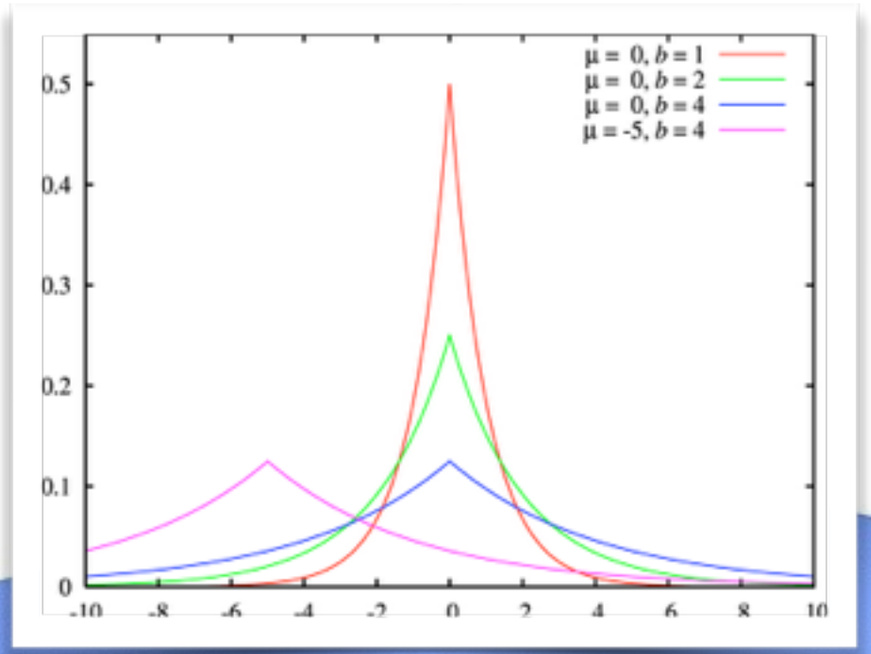
Add random noise

Differential Privacy

K provides ϵ-differential privacy if:

$$\mathcal{D}_p(\mathbf{K(x)}, \mathbf{K(x')}) \leq \epsilon \qquad \forall\ x \sim x'$$

- $x \sim x'$ if they differ in just one entry.
- Same answers with **similar** probabilities.
- Is **independent from the attacker's knowledge**.
- We usually add noise from a **Laplace** distribution.

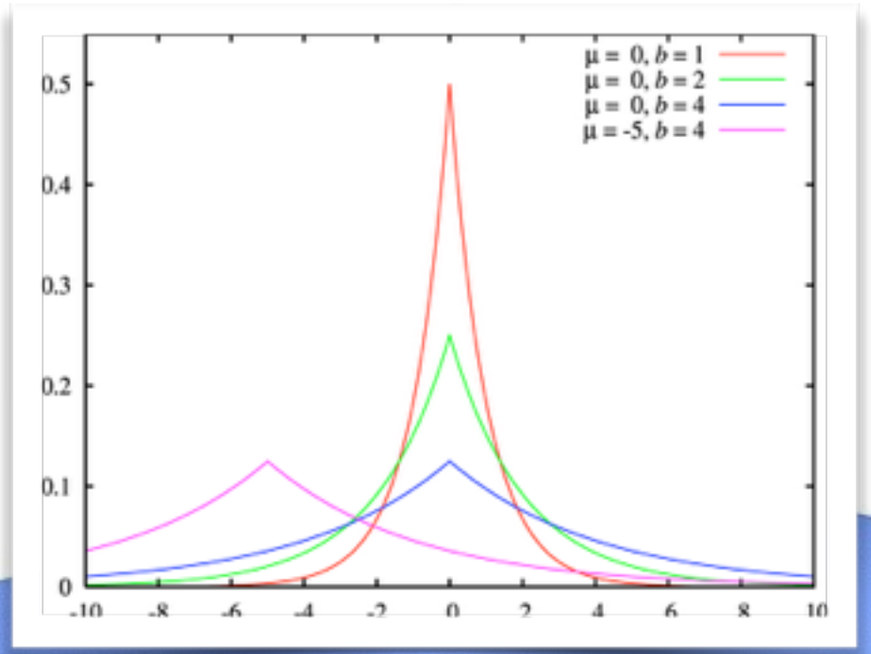


Differential Privacy

K provides ϵ-differential privacy if:

$$\mathcal{D}_{\mathcal{P}}(K(x), K(x')) \leq \epsilon\, d_h(x,x') \qquad \forall\ x,\ x'$$

- x~x′ if they differ in just one entry.
- Same answers with **similar** probabilities.
- Is **independent from the attacker's knowledge**.
- We usually add noise from a **Laplace** distribution.



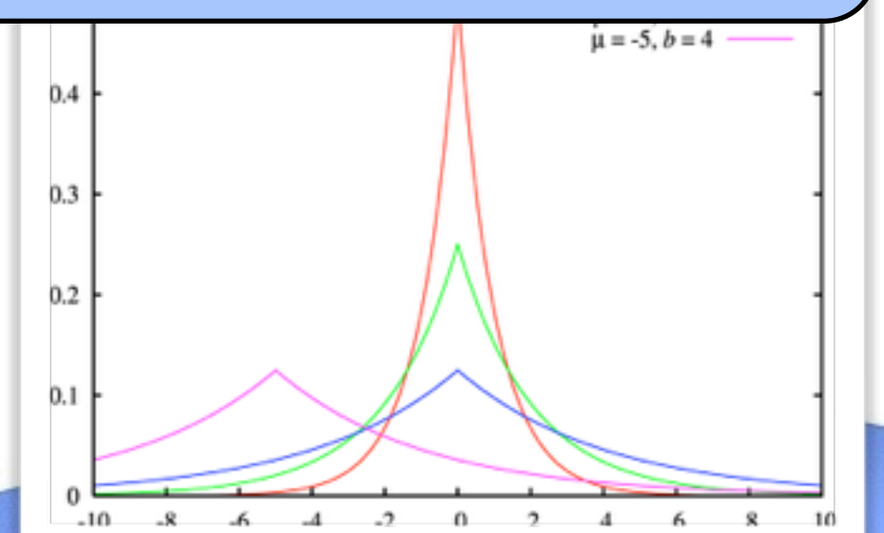
Differential Privacy

K provides ϵ-differential privacy if:

$$\mathcal{D}_{\mathcal{P}}(K(x), K(x')) \leq \epsilon\, d_h(x,x') \qquad \forall\ x,\ x'$$

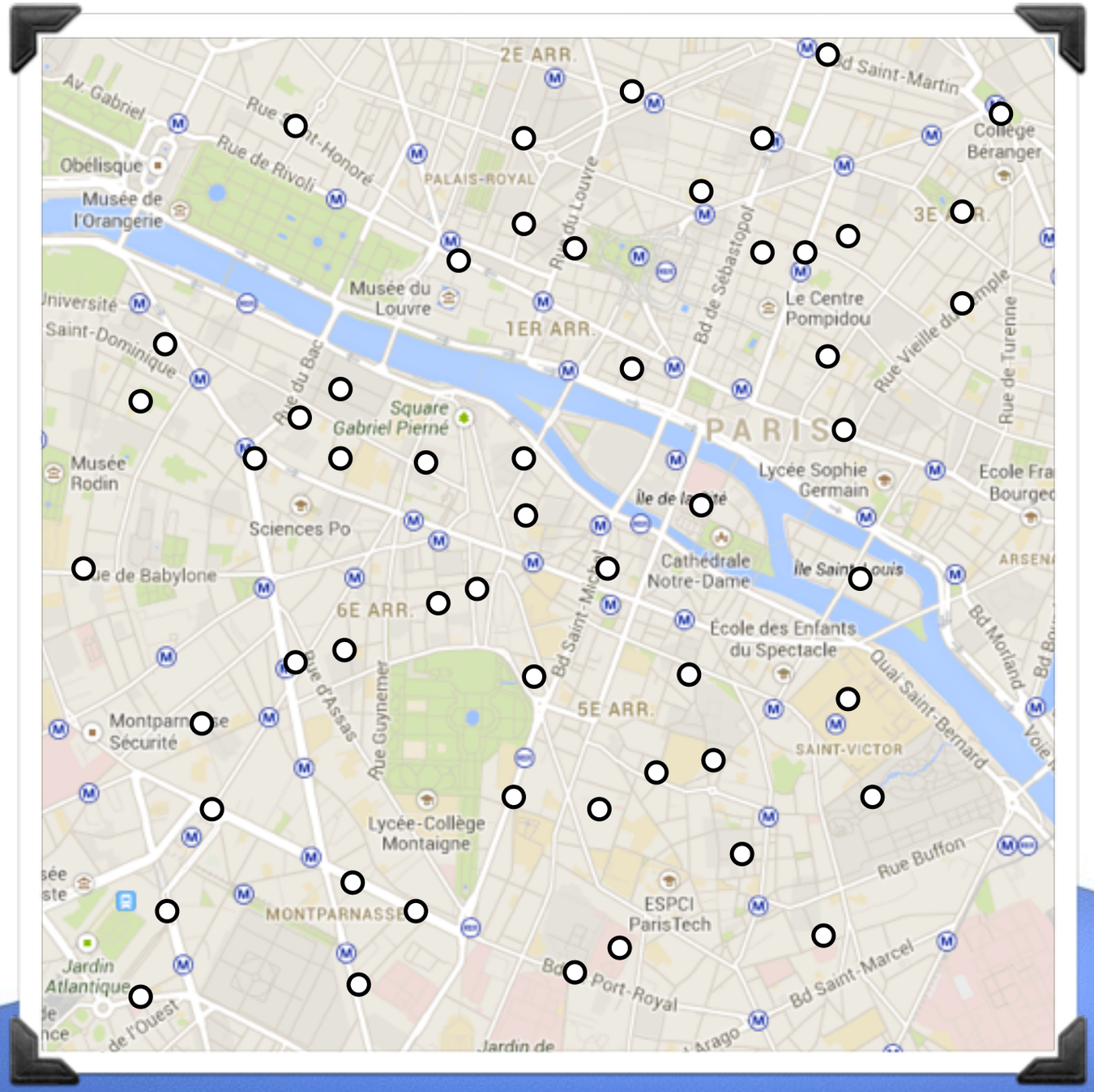
distinguishability level between x and x'

- x~x′ if they differ in just one entry.
- Same answers with **similar** probabilities.
- Is **independent from the attacker's knowledge**.
- We usually add noise from a **Laplace** distribution.



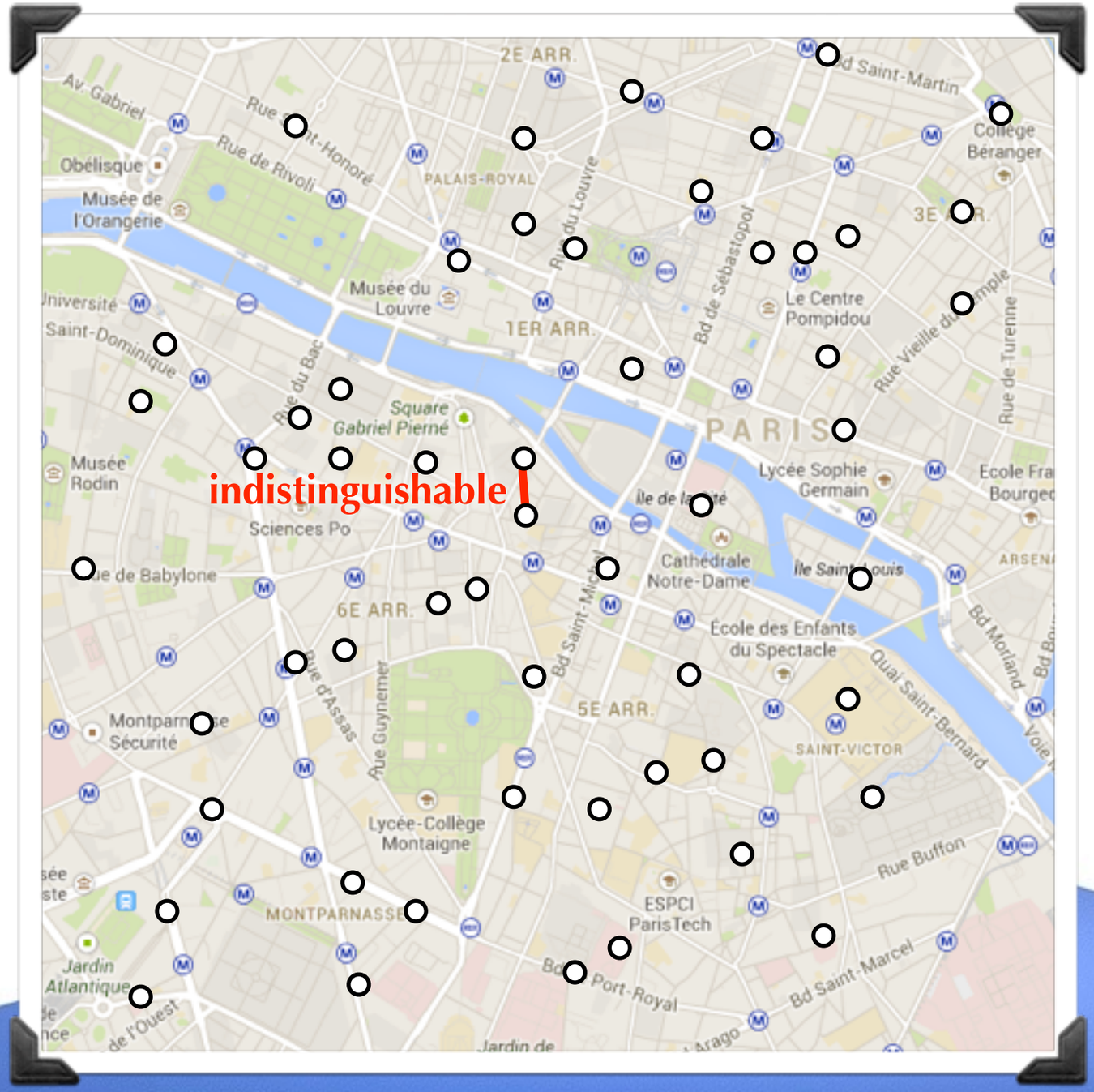
Towards a Definition

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



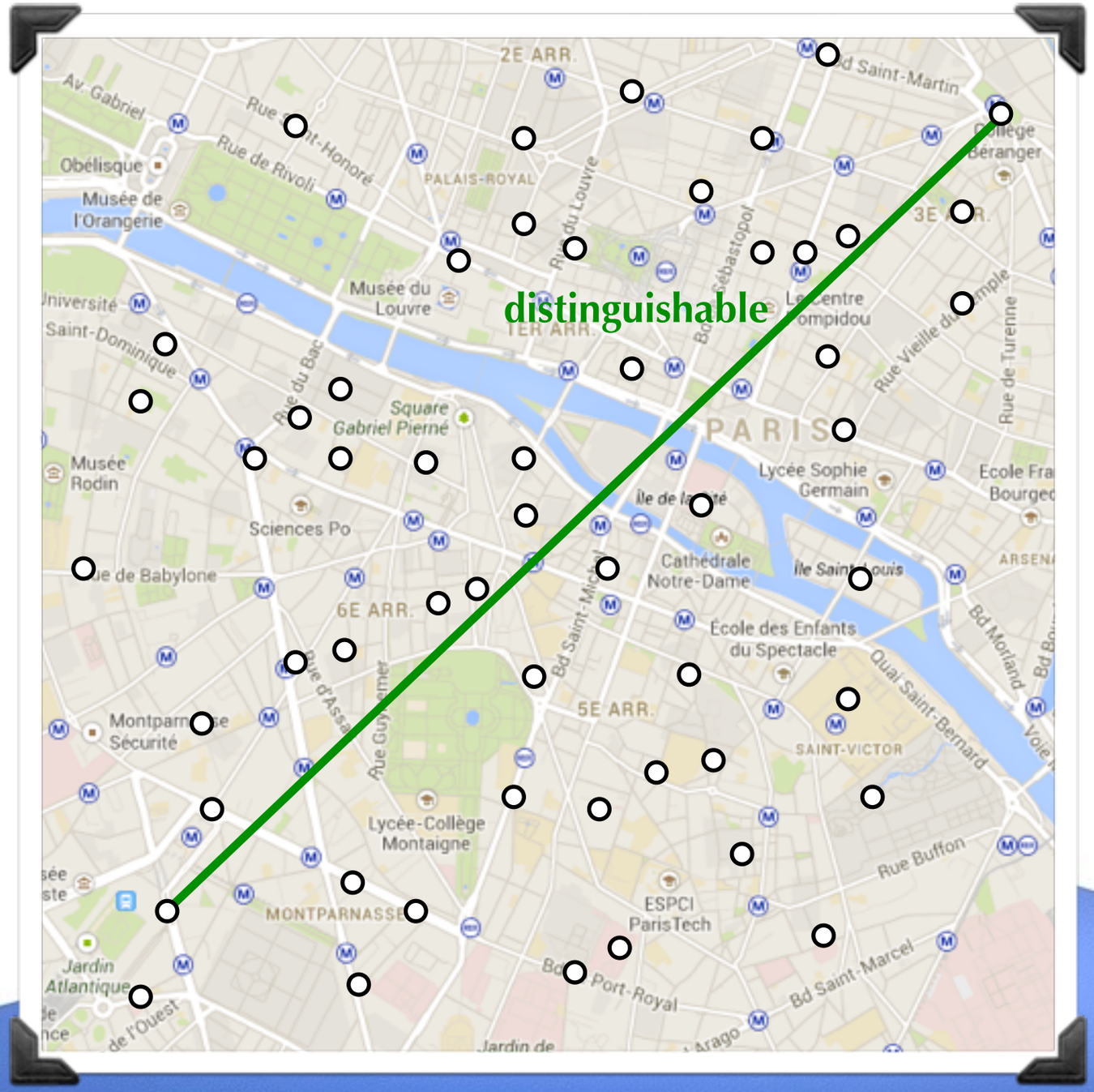
Towards a Definition

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



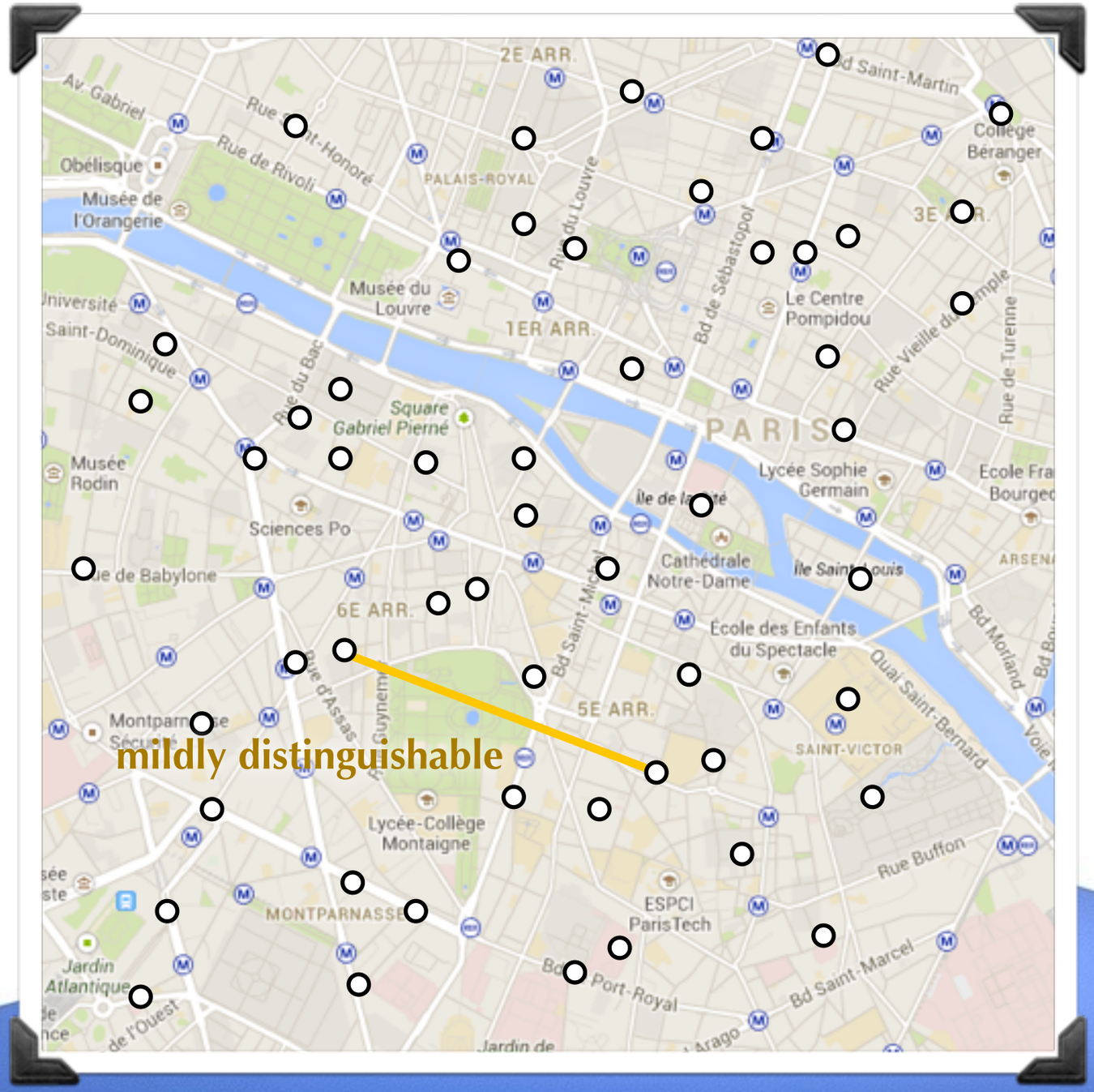
Towards a Definition

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



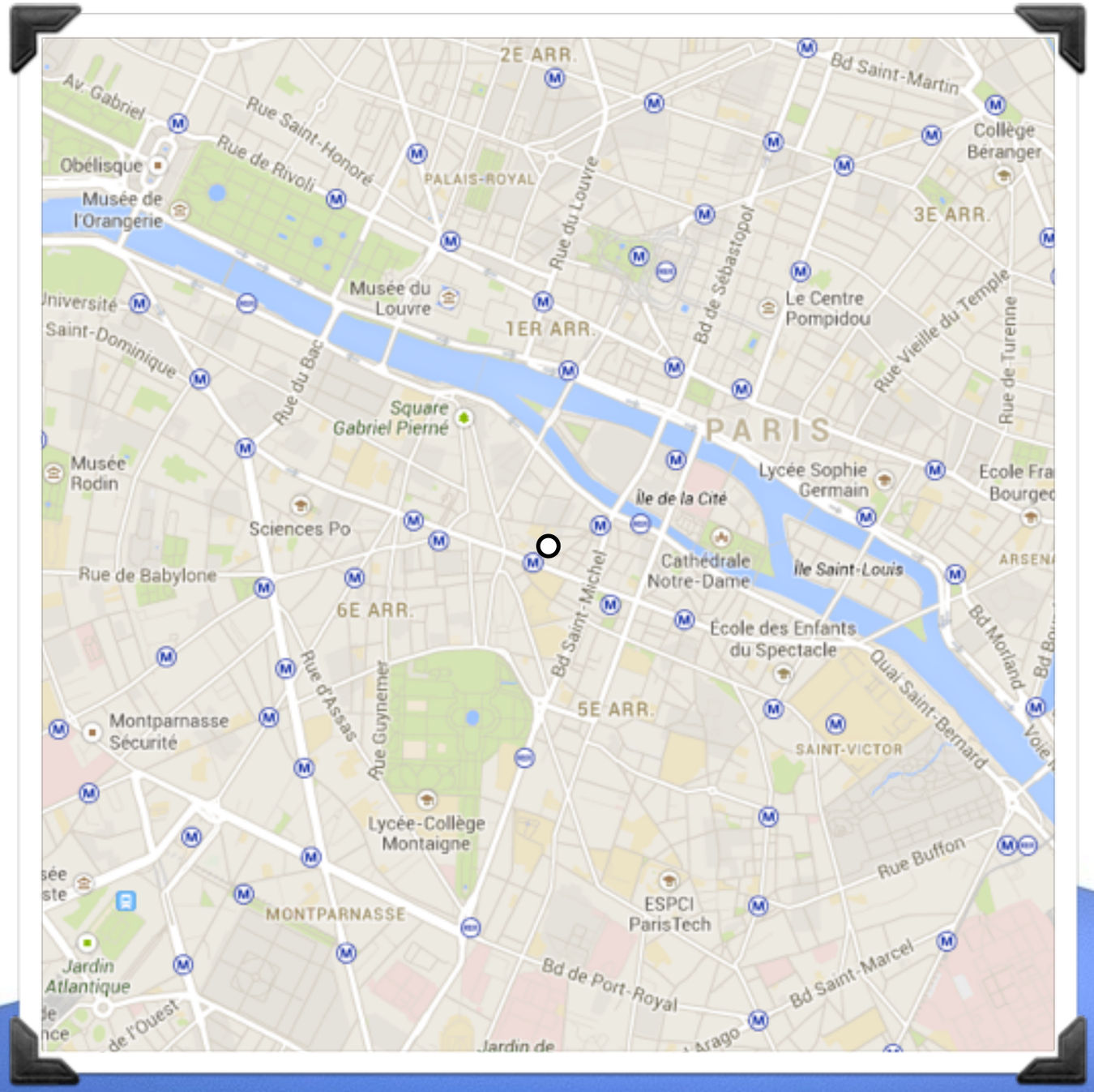
Towards a Definition

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



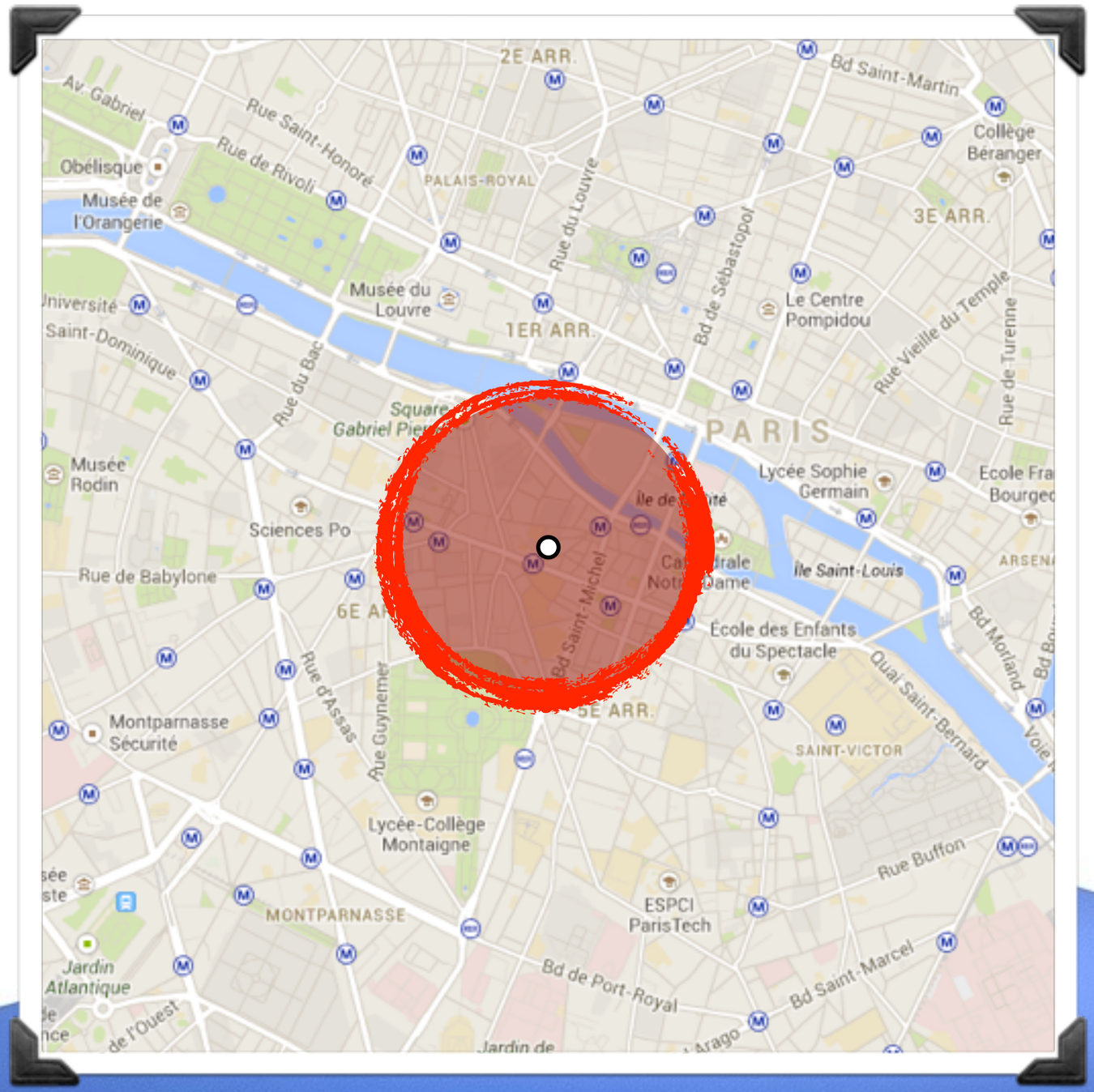
Towards a Definition

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



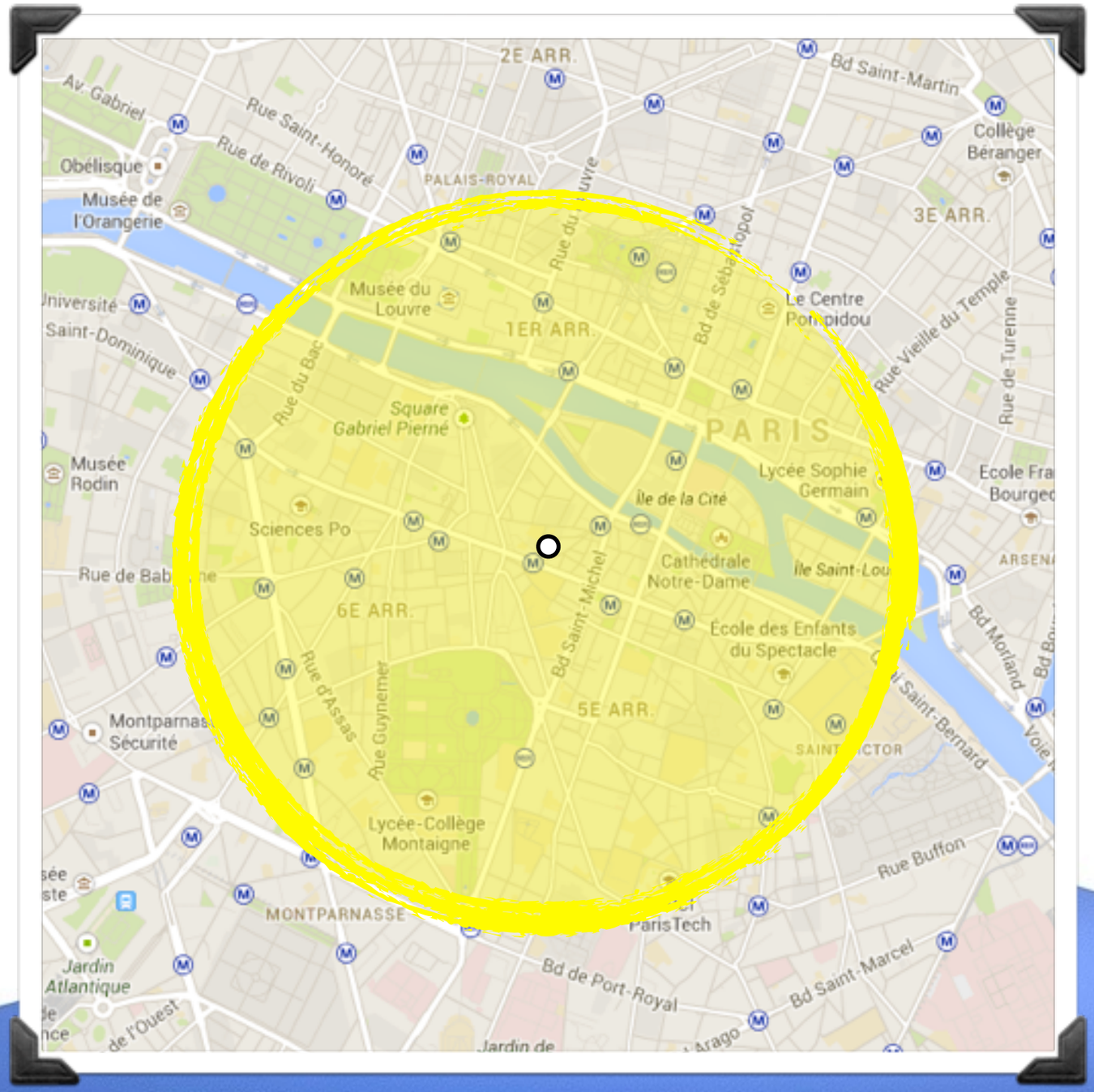
Towards a Definition

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



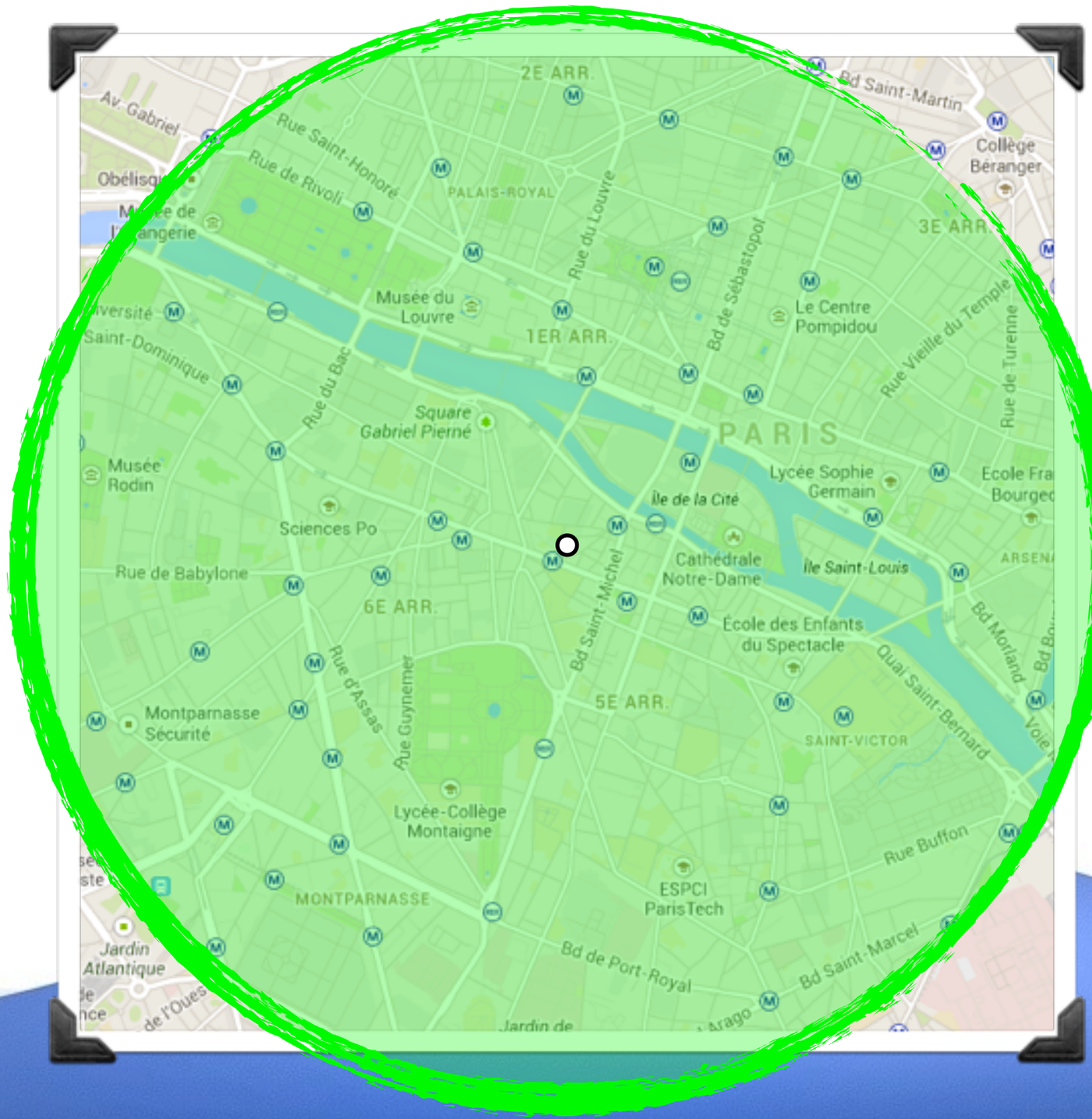
Towards a Definition

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



Towards a Definition

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



d_X-privacy

- We can consider the **set X of possible locations** as the set of secrets, with corresponding metric d_X (e.g. the **Euclidian distance**).

A location obfuscation mechanism K provides **ϵd_X-privacy** if:

$$\mathcal{D}_{\mathcal{P}}(K(x), K(x')) \leq \epsilon\, d_X(x,x') \qquad \forall\ x, x'$$

- Instance of a generalized version of differential privacy.
- Proposed before (e.g. [Pierce et al., ICFP 2010]).
- Deeply studied by us [Chatzikokolakis et al., PETS 2013].

d_X-privacy

Broadening the scope of differential privacy using metrics. PETS 2013.

Develop the general theory of d_X-privacy.

- Characterization results.
- Optimality results.

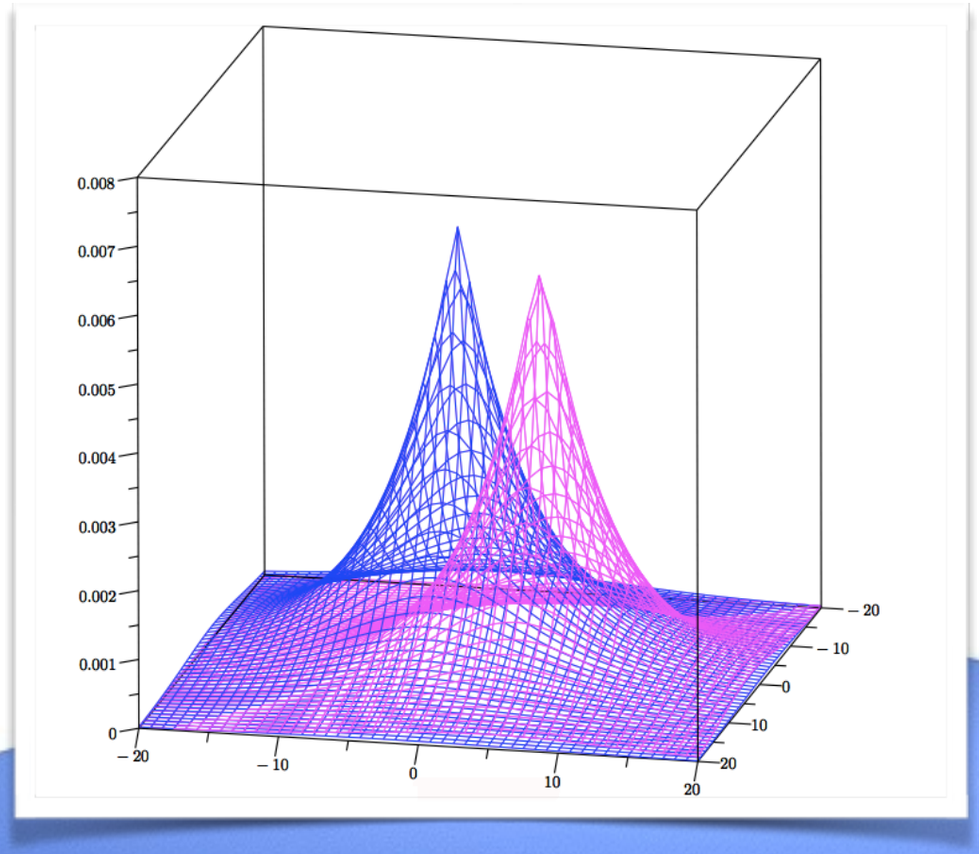
What can we achieve with metrics?

- Strengthen differential privacy.
- Protect the accuracy, and not the complete value.
- Different contexts, not only databases.

The Planar Laplacian Mechanism

- Add noise from a **2-dimensional Laplace distribution**.
- Not the same as applying one-dimensional Laplace noise to each coordinate.

- Independent from the size of **X**.
- Easy to compute.
- Independent from the **user**.



The d_X-optimal Mechanism

- π : user profile, probability distribution over **X**.
- $\boldsymbol{d_X}$: privacy metric on **X**.
- **SQL**(K, π) : expected error of the mechanism K w.r.t. π

$$\mathbf{SQL}(K, \pi) = \sum_{x,z \in \mathbf{x}} \pi(x)\, k_{xz}\, d_X(x,z)$$

Objective: a d_X-optimal mechanism K, meaning

- K is **ϵd_X-private**.
- SQL(K, π) is **as small as possible**.

The d_X-optimal Mechanism

We get K by solving a linear optimization problem:

Choose: k_{xz} $\forall$ x,z

To minimize: SQL(K, π)

Subject to:

$k_{xz} \leq e^{\epsilon d_X(x,x')} k_{x'z}$ $\forall$ x,x′,z **(d_X-privacy)**

$k_{xz} \geq 0$ $\forall$ x,z

$\sum_{z \in \mathbf{x}} k_{xz} = 1$ $\forall$ x

The d_X-optimal Mechanism

We get K by solving a linear optimization problem:

Choose: k_{xz} $\forall$ x,z

To minimize: SQL(K, π)

Subject to:

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

$k_{xz} \leq e^{\epsilon d_X(x,x')} k_{x'z}$ $\forall$ x,x′,z **(d_X-privacy)**

$k_{xz} \geq 0$ $\forall$ x,z

$\sum_{z \in \mathbf{X}} k_{xz} = 1$ $\forall$ x

Spanners

Idea: approximate distances by using spanners.

A weighed graph **G** = (**X**, E, w) is a **spanner** of **X** if

$$w(x,x') = d_X(x,x') \qquad \forall\ (x,x') \in E$$

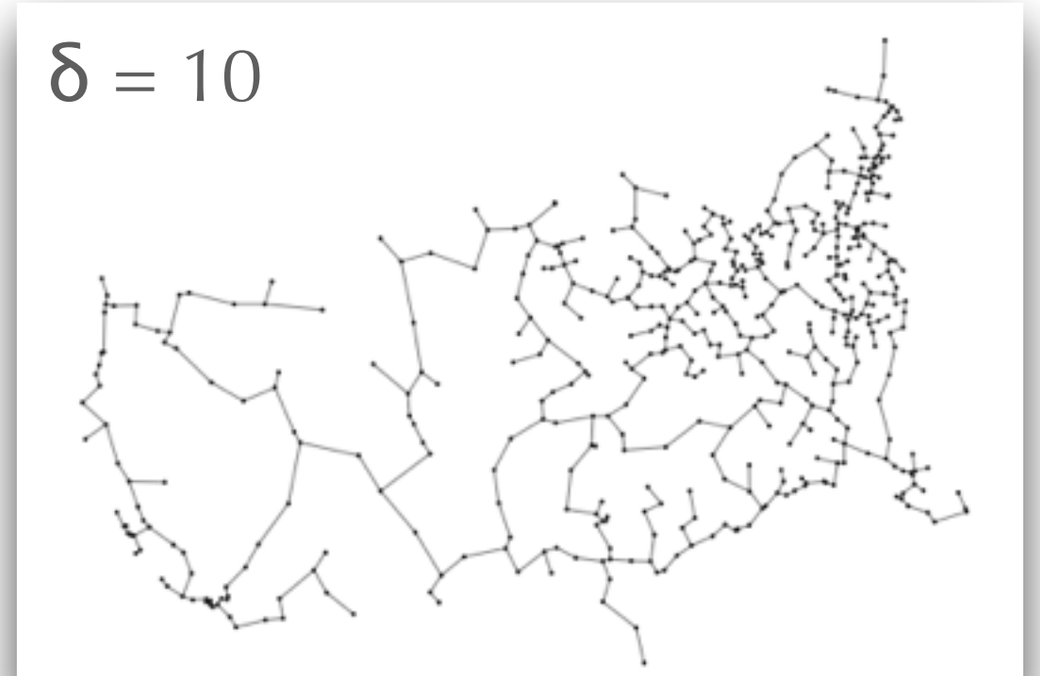
$d_G(x,x')$ is the weight of a minimum path between x and x′ in **G**.

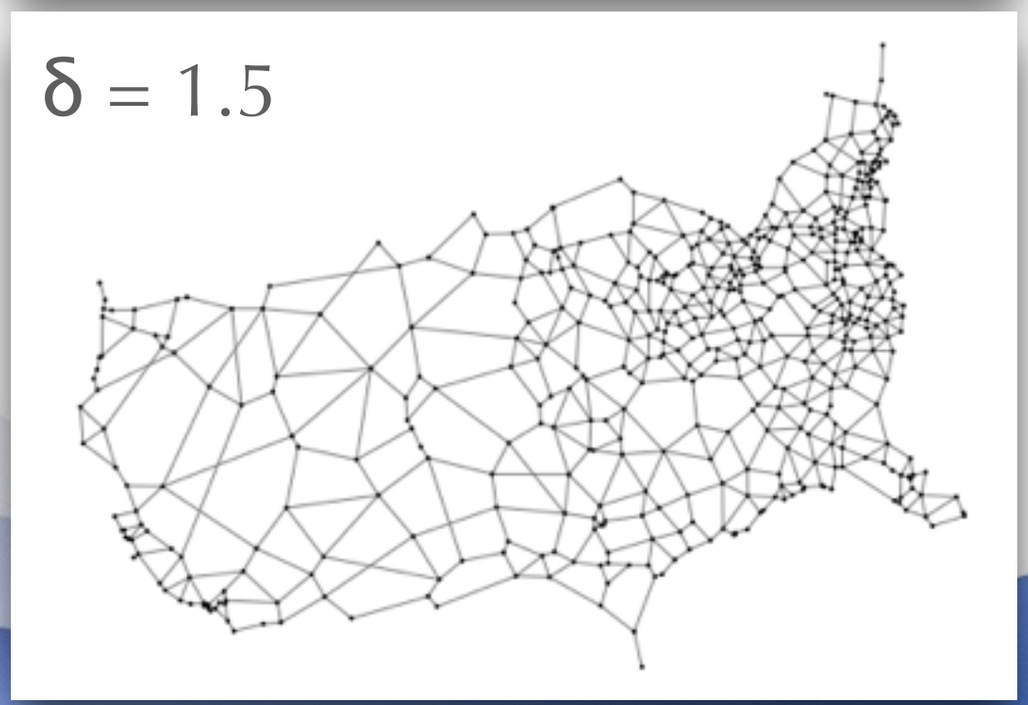
The **dilation** of a spanner **G** is:

$$\delta = \mathbf{max}_{x \neq x'}\ d_G(x,x') \,/\, d_X(x,x')$$

δ is the **worst ratio** between a distance in **G** and the corresponding real one.

Spanners





Spanners

d_G is an **approximation** of d_X.

Fact: If we take d_G as the metric of **X**, we only need to specify the constraints corresponding to edges of **G** in the linear program.

Proposition: If K is **$(\epsilon/\delta)d_G$-private**, then it is **ϵd_X-private.**

We can use d_G to reduce the number of constraints in the optimization problem.

The d_G-optimal Mechanism

Let **G** = (**X**, E, w) be a spanner with dilation **δ**

Choose:	k_{xz}	$\forall$ x,z
To minimize:	SQL(K, π)	

Subject to:

$k_{xz} \leq e^{(\epsilon/\delta) d_G(x,x')} k_{x'z}$	$\forall (x,x') \in E,\ z \in \mathbf{X}$
$k_{xz} \geq 0$	$\forall$ x,z
$\sum_{z \in \mathbf{X}} k_{xz} = 1$	$\forall$ x

The d_G-optimal Mechanism

Let $\mathbf{G} = (\mathbf{X}, E, w)$ be a spanner with dilation δ

Choose: k_{xz} $\quad \forall\ x,z$

To minimize: $SQL(K, \pi)$

Subject to:

$|E|\cdot|\mathbf{X}|$ constraints

$k_{xz} \leq e^{(\epsilon/\delta)d_G(x,x')}\, k_{x'z}$ $\quad \forall\ (x,x') \in E,\ z \in \mathbf{X}$

$k_{xz} \geq 0$ $\quad \forall\ x,z$

$\sum_{z \in \mathbf{X}} k_{xz} = 1$ $\quad \forall\ x$

The d_G-optimal Mechanism

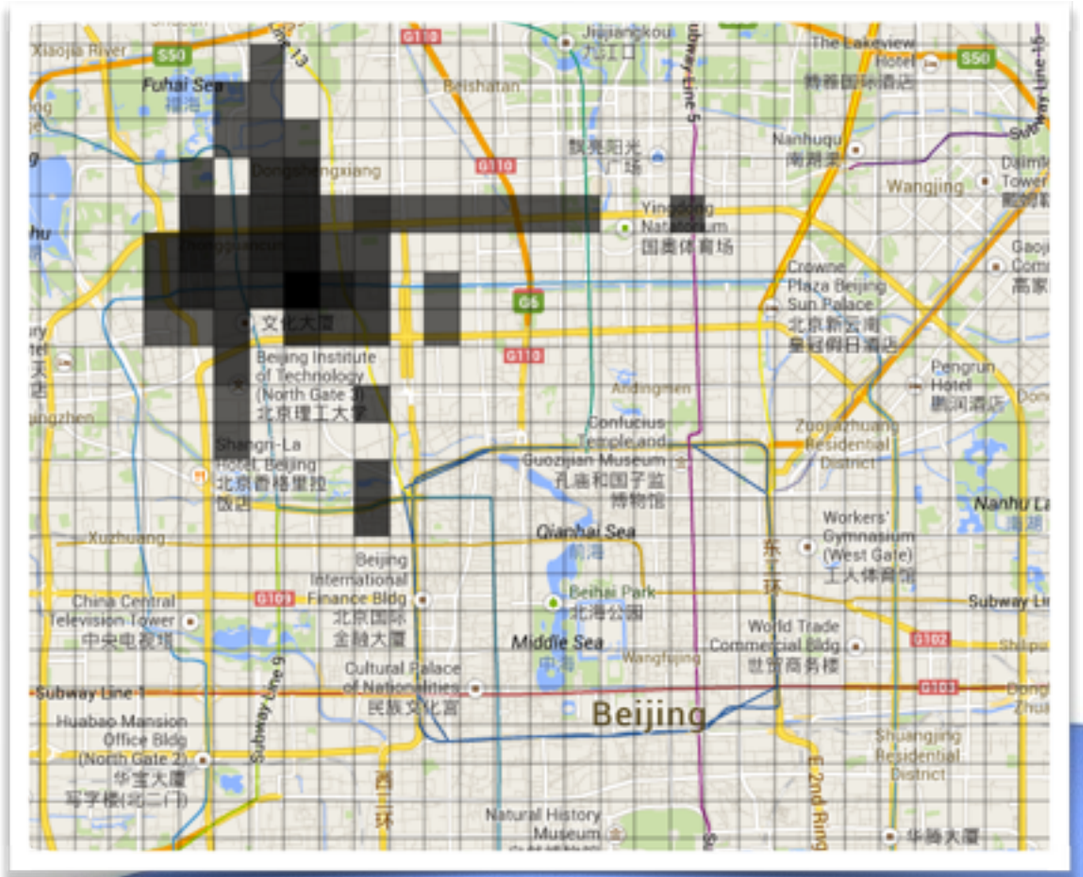
- A d_G-optimal mechanism is not necessarily d_X-optimal.

- Smaller δ implies smaller SQL, but higher amount of constraints.

- If $|E| = O(|\mathbf{X}|)$, then the amount of constraints is $O(|\mathbf{X}|^2)$.

Evaluation

- Traces from the **GeoLife** GPS Trajectories dataset.
- Division of the map into a grid, considering the 50 most popular regions.
- **X** = centres of these regions.
- d_X = Euclidean distance



Evaluation

- The user profile π is constructed using the traces of one user.
- Three mechanisms considered:
  1. The d_G-optimal mechanism with $\delta = 1.1$ under π (**DPOpt**).
  2. The mechanism by Shokri et al., also under π (**EEOpt**).
  3. A discretized version of the Planar Laplacian mechanism (**PL**).
- All mechanism were configured to **the same SQL**.

Evaluation

Privacy measure: **expected error of the adversary**.

$$\textbf{AdvError}(K) = \sum_{x,x',z \in \mathbf{x}} \pi(x)\, k_{xz}\, h_{zx'}\, d_X(x,x')$$

Evaluation

Privacy measure: **expected error of the adversary**.

$$\textsc{AdvError}(K) = \sum_{x,x',z \in \mathbf{x}} \pi(x)\, k_{xz}\, h_{zx'}\, d_X(x,x')$$

adversary's remapping

Evaluation

Privacy measure: **expected error of the adversary**.

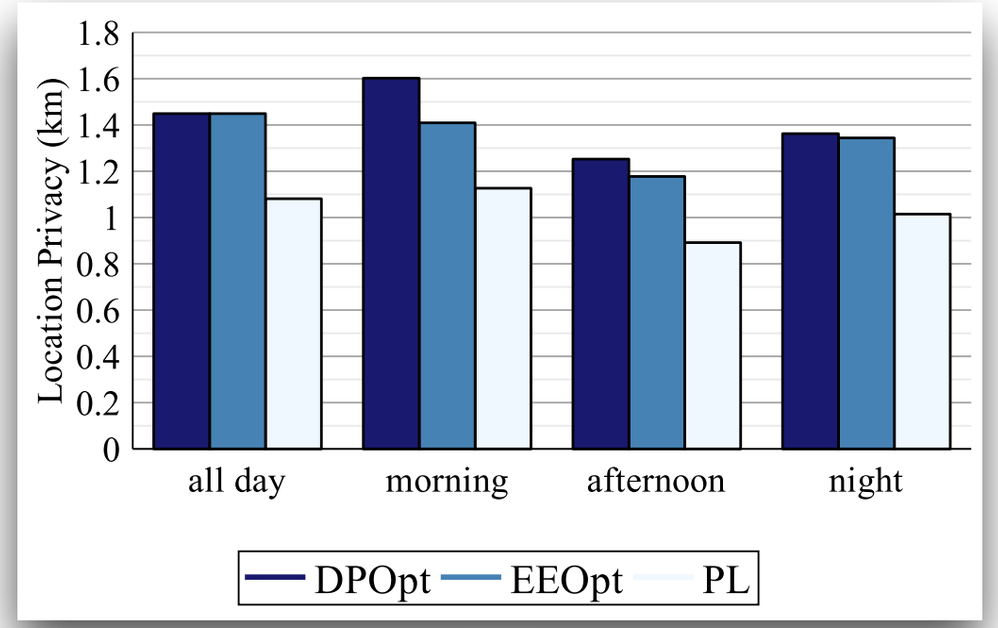
$$\textbf{AdvError}(K) = \sum_{x,x',z \in \mathbf{x}} \pi(x)\, k_{xz}\, h_{zx'}\, d_X(x,x')$$

adversary's remapping

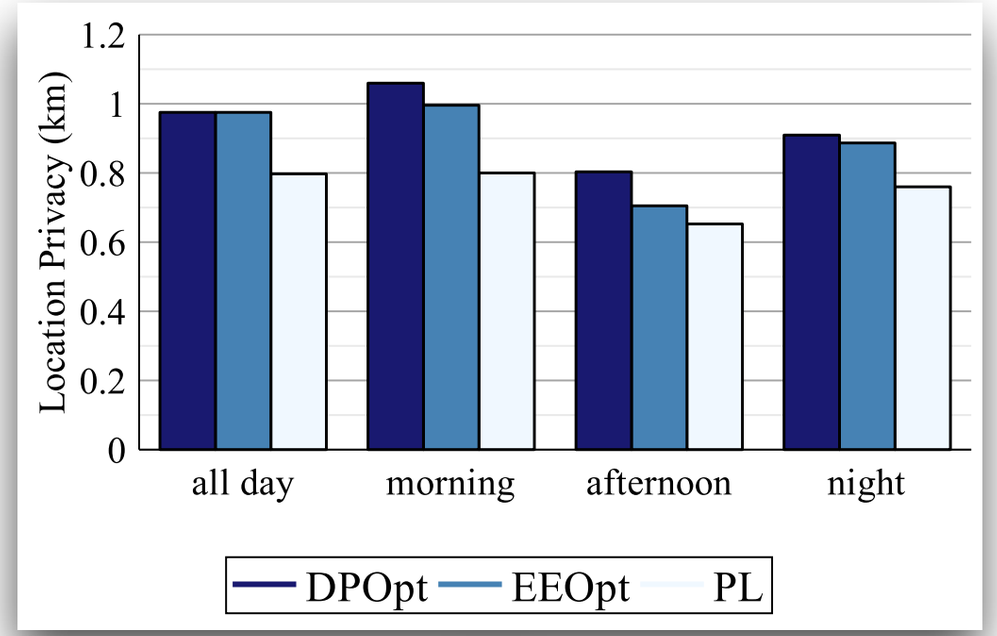
Evaluation **under different priors**.

- All day (the one used to construct the mechanisms).
- Morning (7:00 to 12:00).
- Afternoon (12:00 to 19:00).
- Night (19:00 to 7:00).

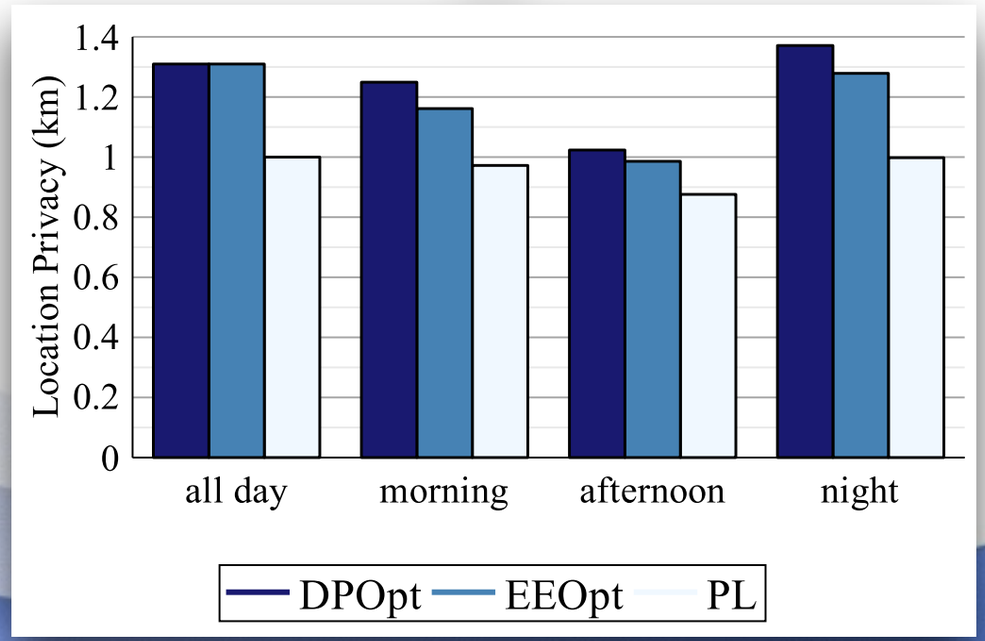
Evaluation



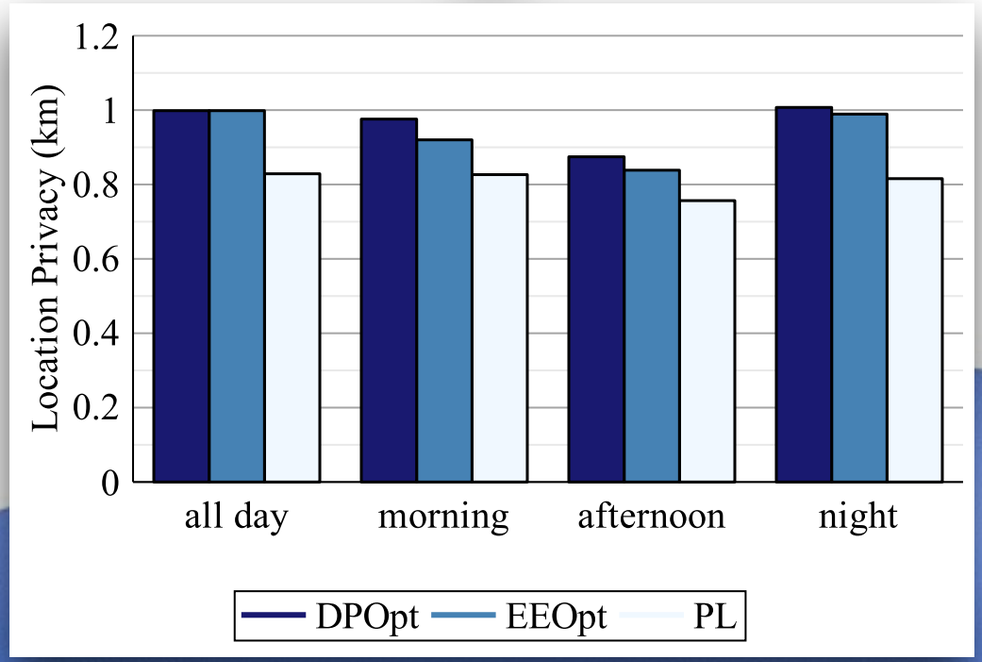
(a)



(b)

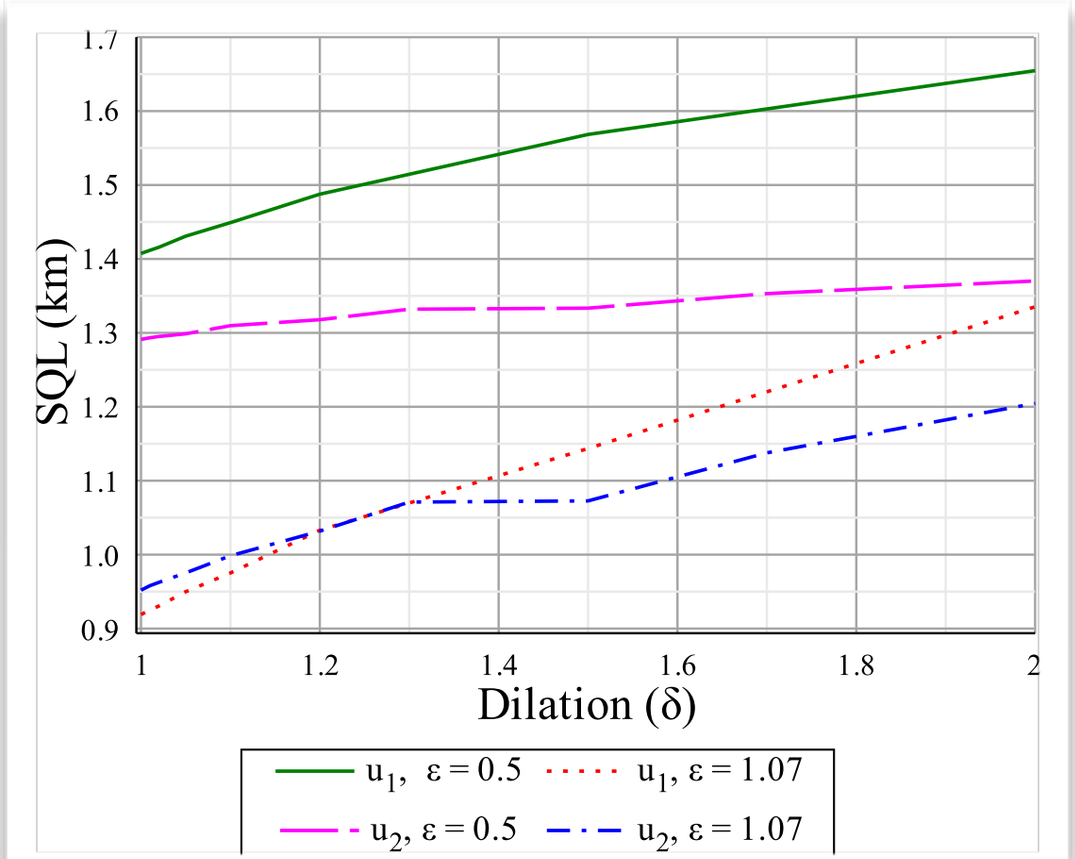


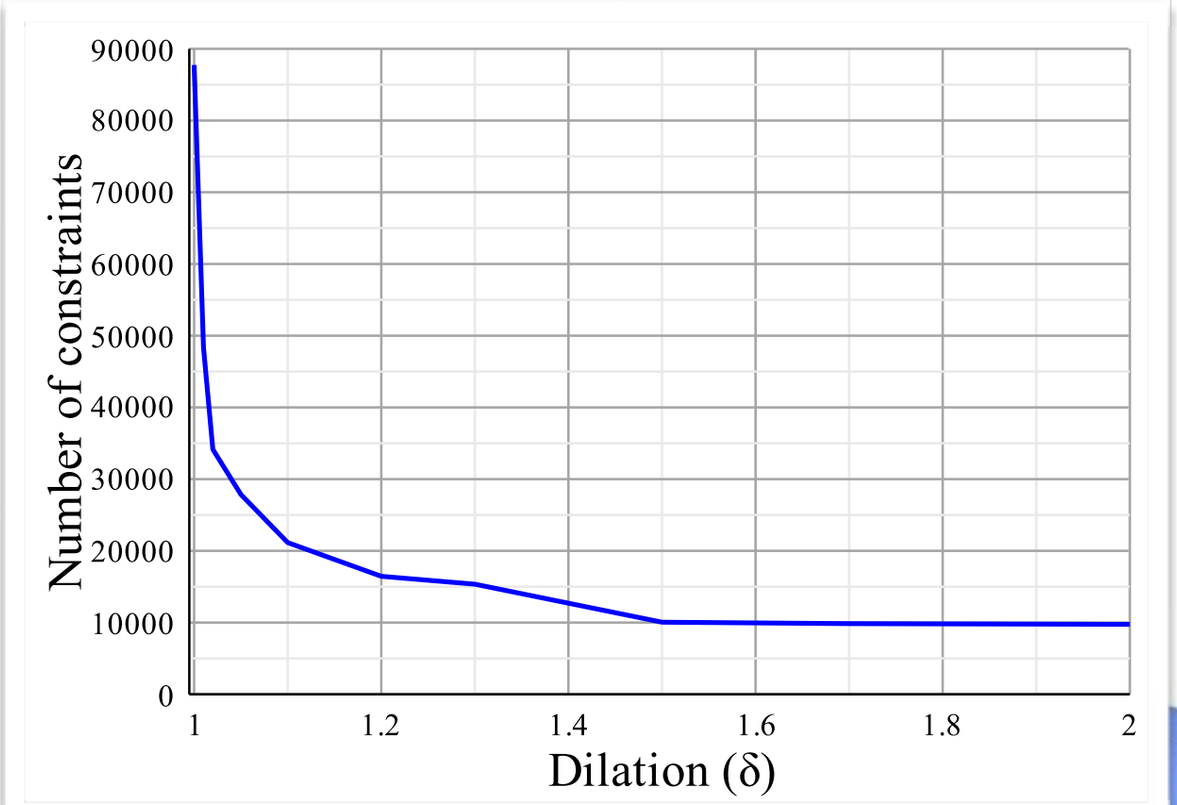
(c)



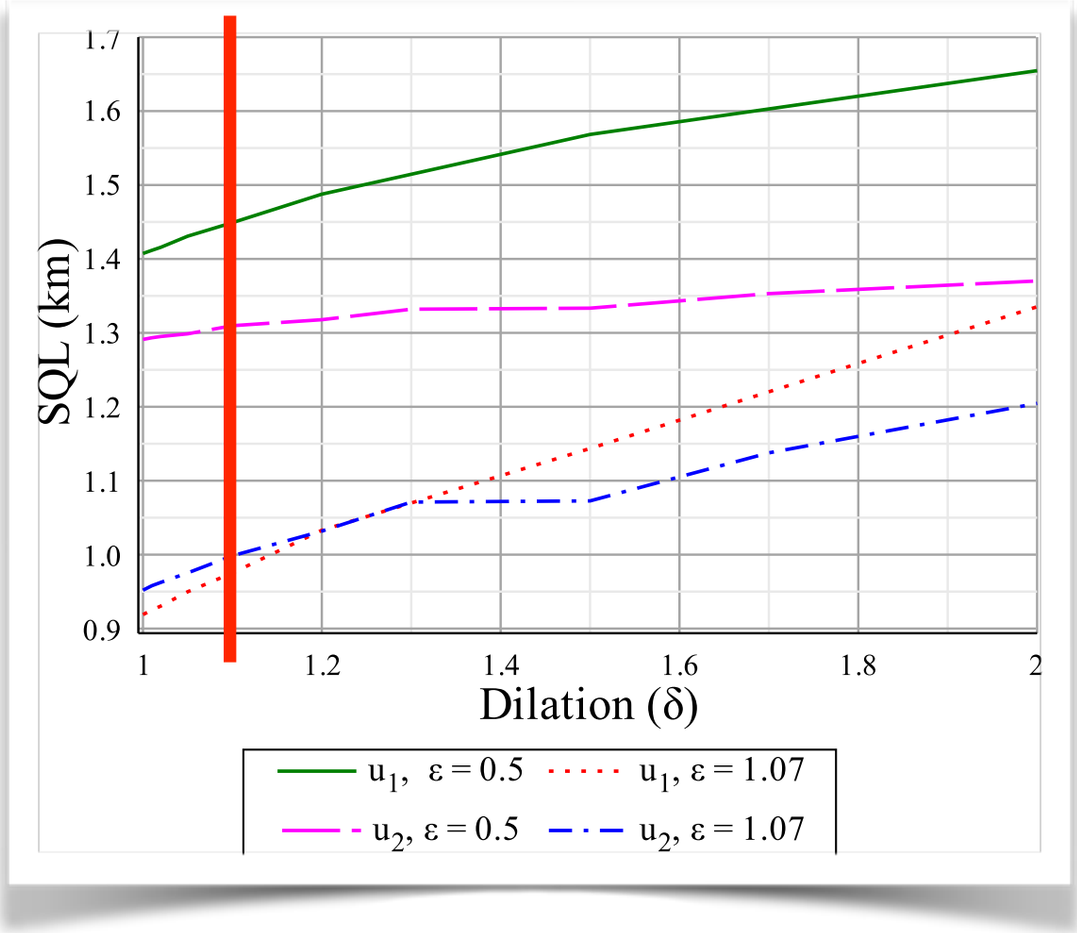
(d)

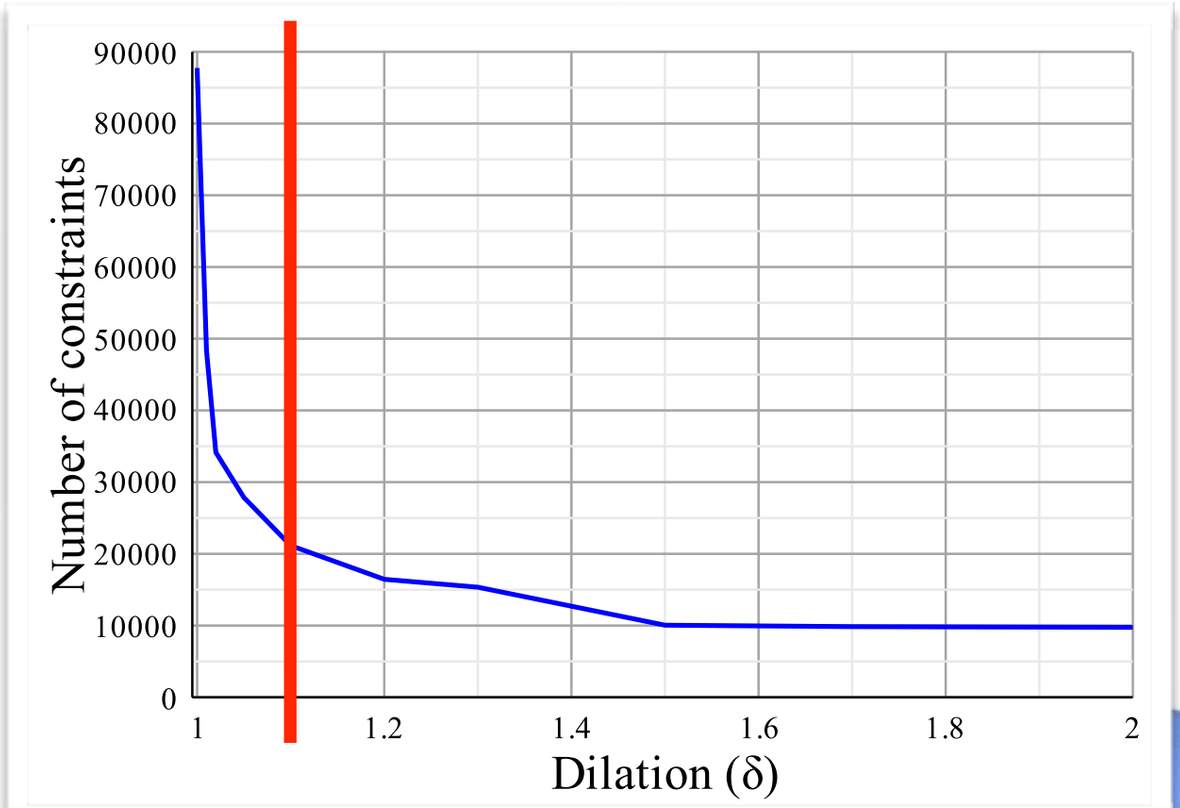
Evaluation


SQL (km)
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
1
1.2
1.4
1.6
1.8
2
Dilation (δ)
u1, ε = 0.5
u1, ε = 1.07
u2, ε = 0.5
u2, ε = 1.07


Number of constraints
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
1
1.2
1.4
1.6
1.8
2
Dilation (δ)

Evaluation


SQL (km)
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
1
1.2
1.4
1.6
1.8
2
Dilation (δ)
u1, ε = 0.5
u1, ε = 1.07
u2, ε = 0.5
u2, ε = 1.07


Number of constraints
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
1
1.2
1.4
1.6
1.8
2
Dilation (δ)

Summary

d_X-privacy:

- A natural extension of differential privacy.
- Metrics allow to capture new privacy notions.
- Extend to different contexts, like **location privacy**.

d_X-optimal mechanism:

- Satisfies ϵd_X-privacy (independent from the prior and the user).
- Optimal utility for a given user.
- An approximate mechanism can be obtained more efficiently.
- Performs well with respect to other state-of-the-art mechanism.

Thanks for your attention!

Questions?

“Location Guard” for Chrome