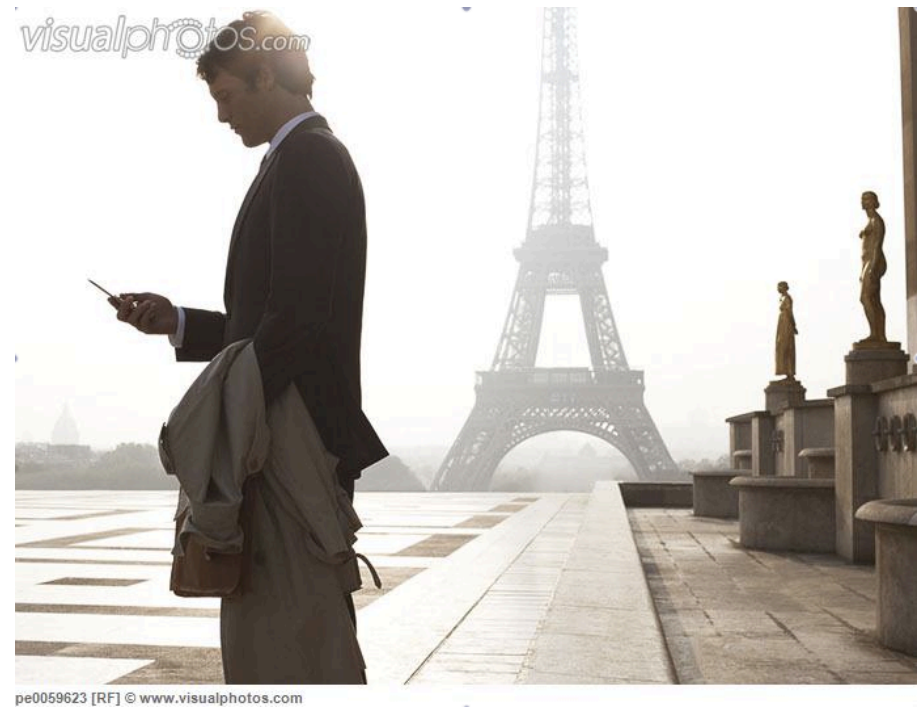
Location privacy via Geo-indistinguishability

Catuscia Palamidessi
INRIA & Ecole Polytechnique

In collaboration with:
Miguel Andres, Nicolas Bordenabe,
Kostas Chatzikokolakis, Marco Stronati

Location Privacy

- Example: use an LBS to find points of interest (restaurants, shops, etc.)
- Revealing the exact location may be dangerous: profiling, inference of sensitive information, etc.
- Revealing an approximate location is usually ok



Geo-indistinguishability

- We do not rely on a trusted third party
- We do not rely on the presence of other users in the proximity
- The notion of geo-indistinguishability is based on an suitable extension of **differential privacy**

Original definition of Differential Privacy (on Databases)

Definition [Dwork et al., 2006]: a randomized mechanism $\mathcal{K}$ provides ε-differential privacy if for all databases x, x' **which are adjacent** (i.e., differ for only one record), and for all $z \in \mathcal{Z}$, we have

$$\frac{p(K = z|X = x)}{p(K = z|X = x')} \leq e^{\epsilon}$$

By the Bayes theorem, this definition corresponds to say that the answer given by K does not change significantly the knowledge about an individual (prior and posterior are close)

Important properties:

- DP is robust with respect to composition of queries: the level of privacy ε decreases linearly with the number of queries
- The definition of DP is independent from the prior

Typical implementation of differential privacy: add Laplacian noise

- Randomized mechanism for a query $f: \mathcal{X} \to \mathcal{Y}$.
- **Add Laplacian noise.** If the exact answer is y, the reported answer is z, with a probability density function defined as:

$$dP_y(z) = c\, e^{-\frac{|z-y|}{\Delta f}\varepsilon}$$

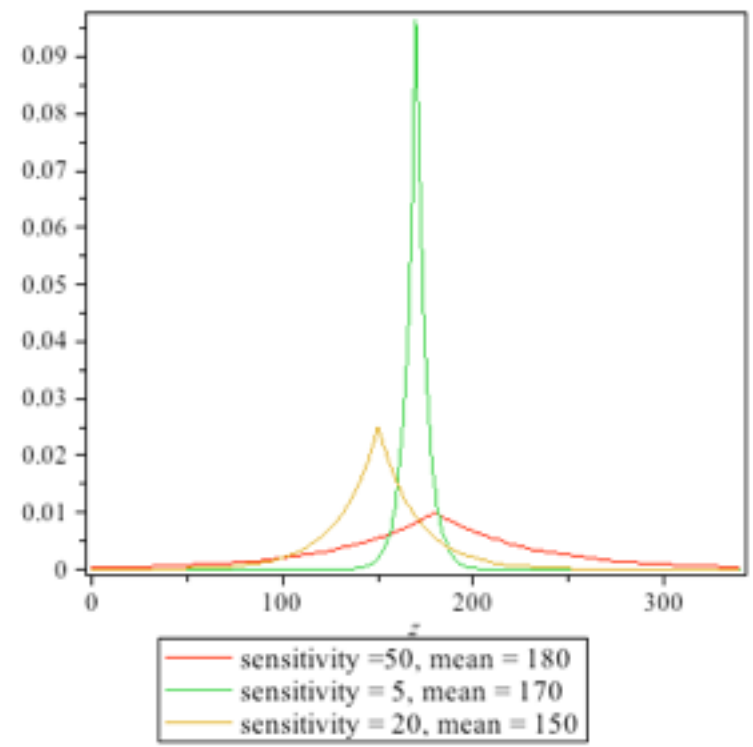
where Δf is the *sensitivity* of f:

$$\Delta f = \max_{x \sim x' \in \mathcal{X}} |f(x) - f(x')|$$

($x \sim x'$ means x and x' are adjacent, i.e., they differ only for one record)

and c is a normalization factor:

$$c = \frac{\varepsilon}{2\,\Delta f}$$



Extending differential privacy to arbitrary metrics

Equivalent definition of DP:

A mechanism is ε-differentially private iff for every pair of databases x, x' and every answer z we have

$$\frac{p(z \mid x)}{p(z \mid x')} \leq e^{\varepsilon\, d_H(x,x')}$$

where d_H is the Hamming distance between x and x', i.e., the number of records in which x and x' differ

Generalization: d-privacy

On a generic domain $\mathcal{X}$ provided with a distance d:

$$\forall x, x' \in \mathcal{X}, \forall z \quad \frac{p(z \mid x)}{p(z \mid x')} \leq e^{\varepsilon\, d(x,x')}$$

Protection of the **accuracy** of the information

Properties of d-privacy :

$$\forall x, x' \in \mathcal{X}, \forall z \quad \frac{p(z \mid x)}{p(z \mid x')} \leq e^{\varepsilon\, d(x,x')}$$

- d-privacy is robust w.r.t. composition: the level of privacy decreases linearly with the number of observations

- d-privacy does not depend on the prior

Location privacy: geo-indistinguishability

d : the Euclidean distance

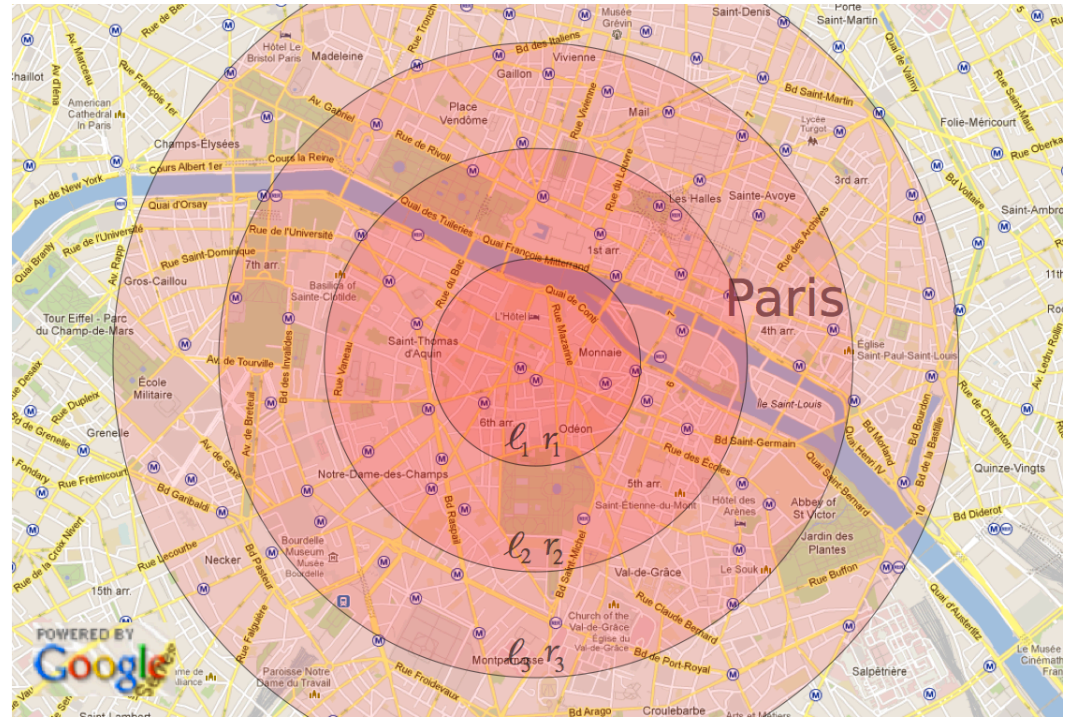
x : the exact location

z : the reported location

$d-$ privacy

$$\frac{p(z|x)}{p(z|x')} \leq e^{\epsilon r}$$

where r is the distance between x and x'



We call this property **geo-indistinguishability**

Note that, since it is a particular case of *d*-privacy, it is, like DP, independent from the prior, and the composition of observations (reported locations) decreases the level of privacy in a linear way.

Meaning of geo-indistinguishability

Using the Bayes theorem, we can give an alternative, and more intuitive, characterization of the geo-indistinguishability property:

According to the Bayes theorem, the conditional probability of z given x can be seen as a transformation from a prior π on x to a posterior P on x given z, :

$$P(x|z) = \frac{p(z|x)\,\pi(x)}{\sum_{x'} p(z|x')\pi(x')}$$

Hence the property of geo-indistinguishability can be rewritten as:

$$\forall \pi.\ \ \frac{P(x|z)}{P(x'|z)} \leq e^{\epsilon d(x,x')}\,\frac{\pi(x)}{\pi(x')}$$

Note that $\frac{P(x|z)}{P(x'|z)}$ depends on π, but, also in this characterization, we can see that the property of geo-indistinguishability is independent from π (since π is quantified universally)

Meaning of geo-indistinguishability

$$\forall \pi. \quad \frac{P(x|z)}{P(x'|z)} \leq e^{\epsilon d(x,x')} \frac{\pi(x)}{\pi(x')}$$

The closer two points are,
the more they are indistinguishable

The level of distinguishability
also depends on the prior

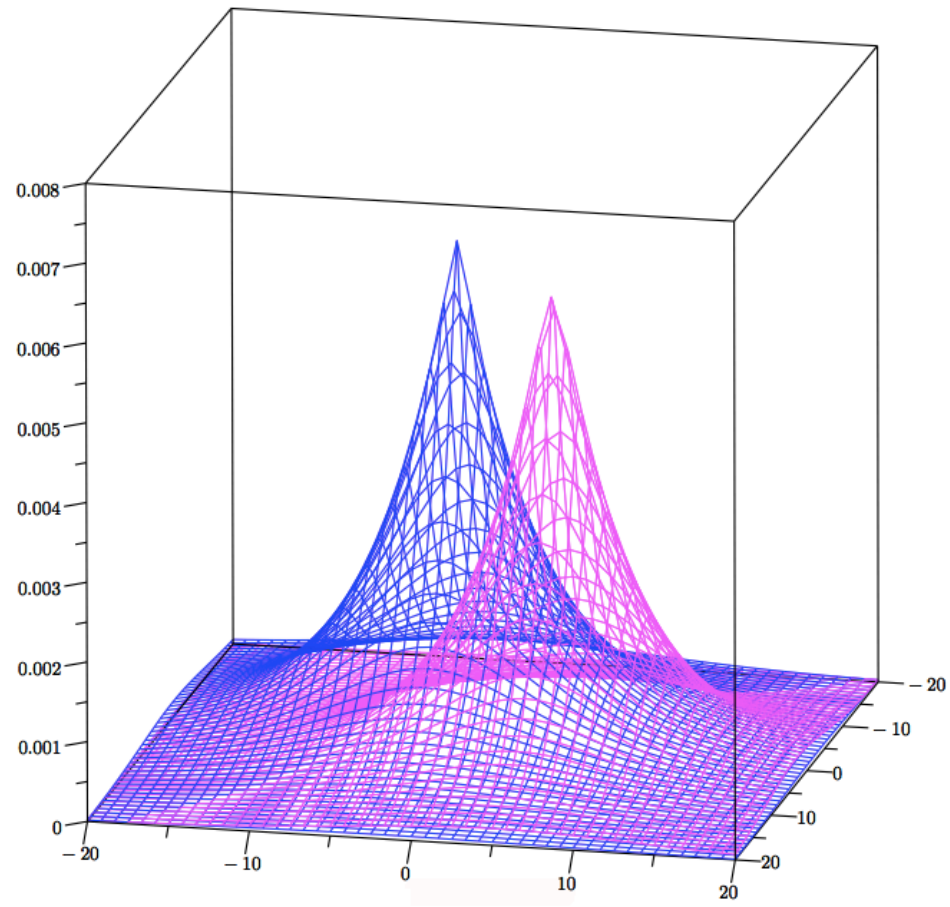
We want to be unable to tell whether the user is in rue Pigalle or at Notre Dame, but it is ok to disclose that he is in Paris and not in London

A *d*-private mechanism for LBS:
Planar laplacian

Bivariate Laplacian

$$dp_x(z) = \frac{\epsilon^2}{2\pi} e^{\epsilon\, d(x,z)}$$

- We have an efficient method to draw points based on polar coordinates
- Afterwards we translate from polar coordinates to standard (latitude, longitude) coordinates.
- Some degradation of the privacy level in single precision, but negligeable in double precision.



Privacy versus utility: evaluation

We have compared the trade off utility-privacy of our mechanism (Planar laplacian) with three other mechanisms in the literature:

- The Optimal Mechanism by Shokri, Theodorakopoulos, Troncoso, Hubaux, Le Boudec. [Shokri et al. CCS 2012]. Note that this mechanism is prior-dependent: it is specifically generated assuming a certain adversary (with a certain prior knowledge), using linear programming techniques. Our mechanism, in contrast, is prior-independent.

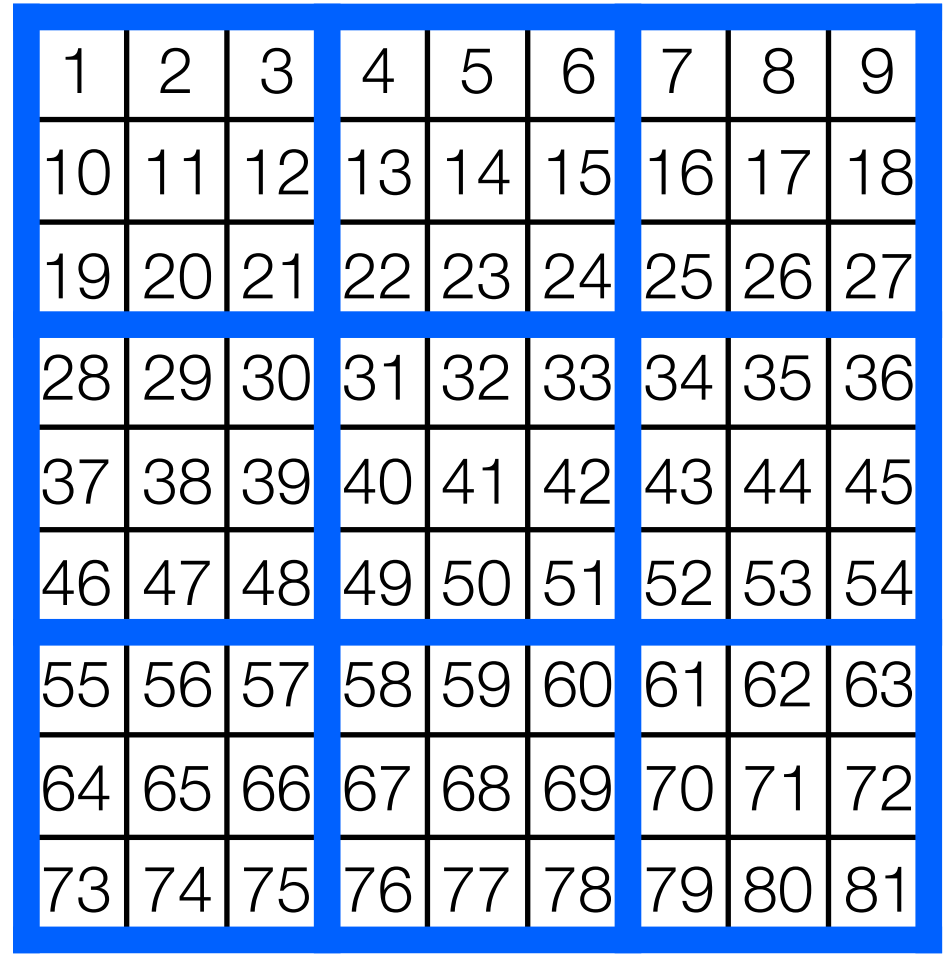
- Two prior-independent mechanisms:

  - Spatial cloacking: We partition the area of interest in zones, and instead of reporting the point, we report the zone in which the point is.

  - The mechanism of Shokri et al., generated assuming uniform prior.

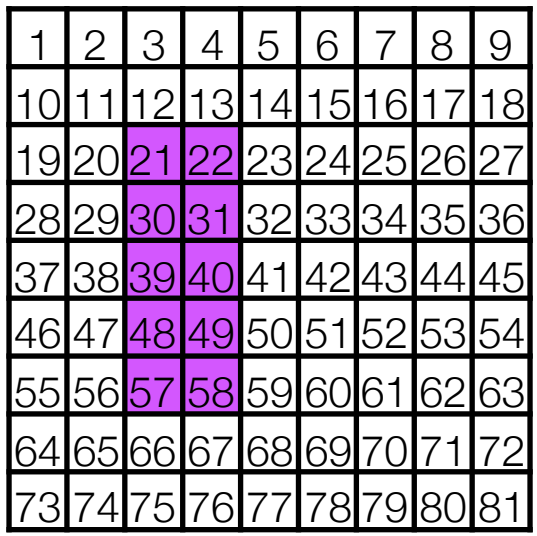
Privacy versus utility: evaluation

- We have designed an ``area of interest'' containing 9x9 = 81 "locations".
- For the cloaking mechanism, we have partitioned the area in 9 zones, indicated by the blue lines

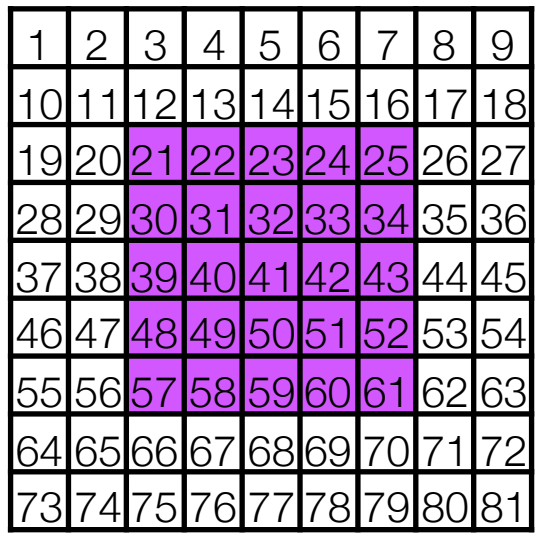


Privacy versus utility: evaluation

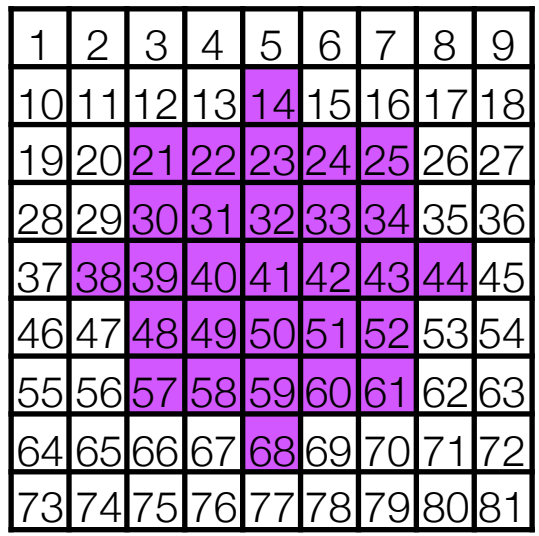
- We configured the four mechanisms so to give the same utility, and we measured their privacy.
- **Utility**: **expected distance** between the true location and the reported one (utility loss) [Shroki et al., CCS 2012]
- **Privacy**: **expected error of the attacker** (using prior information) [Shroki et al., CCS 2012]. Note that we could not use geo-indistinguishability, because our mechanism is the only one that provide geo-indistinguishability
- Priors: concentrated over colored regions



(a)



(b)

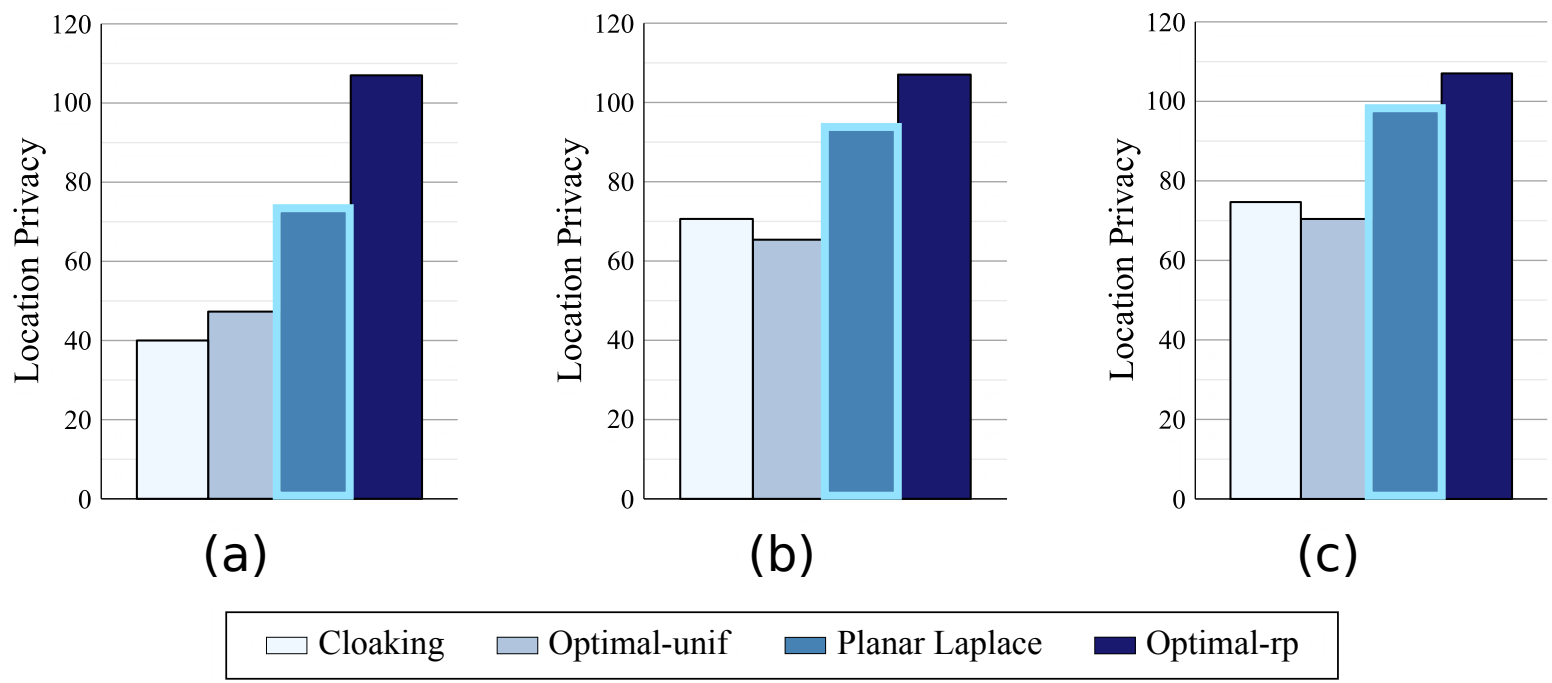


(c)

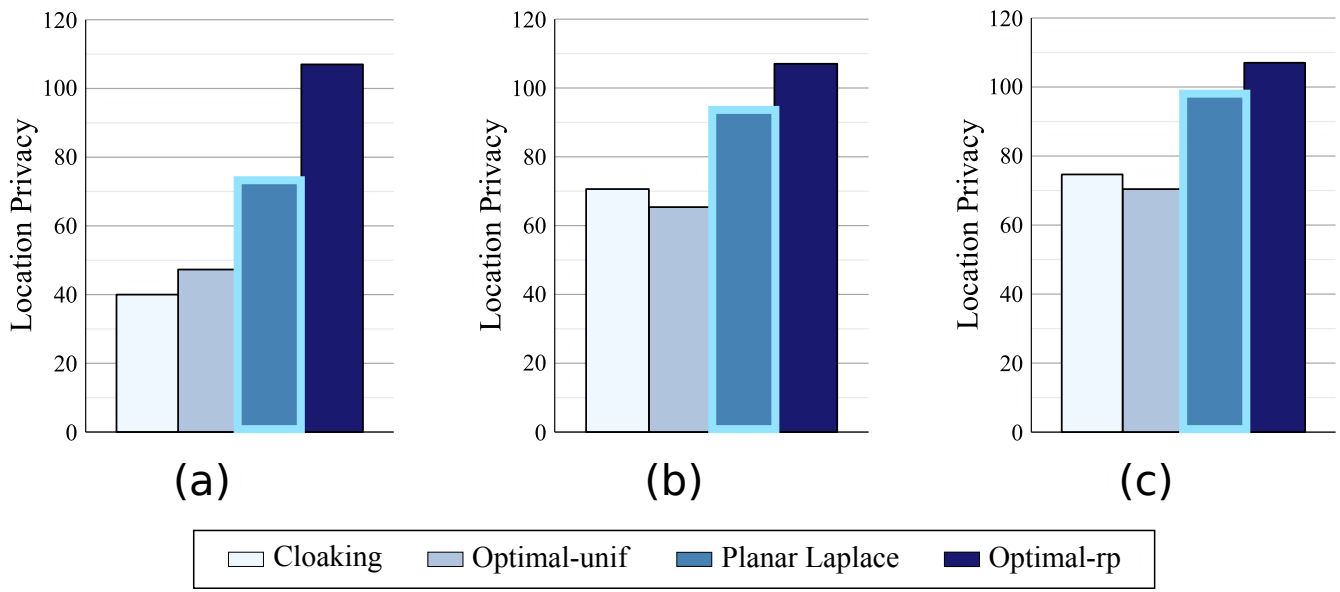
Privacy versus utility: evaluation

The four mechanisms:

- Cloaking,
- Optimal by [Shroki et al. CCS 2012] generated assuming uniform prior
- Ours (Planar Laplacian)
- Optimal by [Shroki et al. CCS 2012] generated assuming the given prior



Privacy versus utility: evaluation

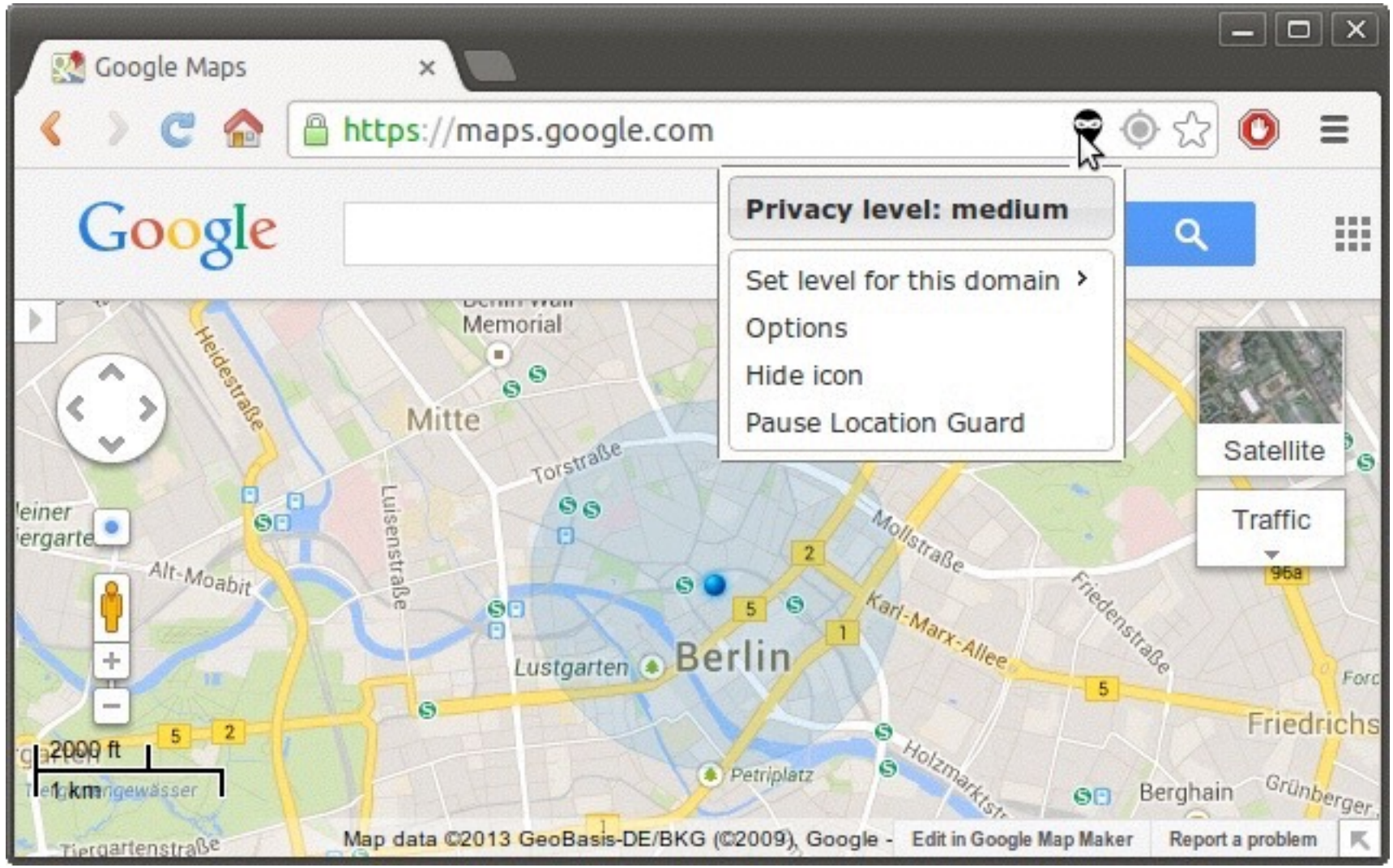


With respect to the privacy measures proposed by [Shokri et al, CCS 2012], our mechanism performs better than the other mechanisms proposed in the literature which are independent from the prior (and therefore from the adversary)

Tool: "Location Guard"

http://www.lix.polytechnique.fr/~kostas/software.html

Extension for Firefox, Chrome, and Opera. It has been released about one year ago, and nowadays it has about 60,000 active users.



Impact

- 60K users of Location Guard

- Nicolas Bordenabe obtained the ACM SIGSAC award for the best PhD thesis in Security and Privacy in 2014

- Collaboration with Renault

Thanks!