Back to the Drawing Board:

Revisiting the Design of Optimal Location Privacy-preserving Mechanisms

Simon Oya, Carmela Troncoso, Fernando Pérez-González

Universidad de Vigo

GPSC
Signal Processing in Communications Group

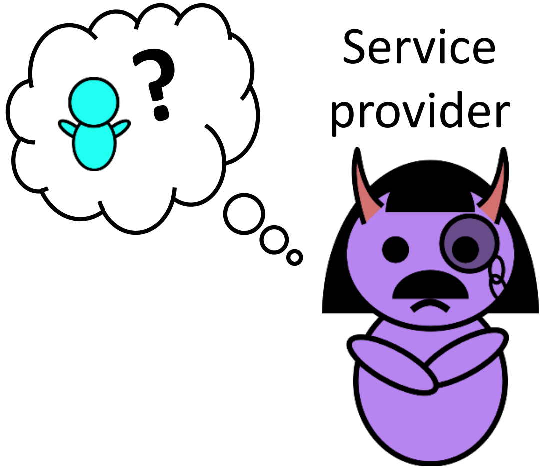
atlanTTic
research center for Telecommunication Technologies

institute IMDEA software

Motivation. Obfuscation-Based Location Privacy.

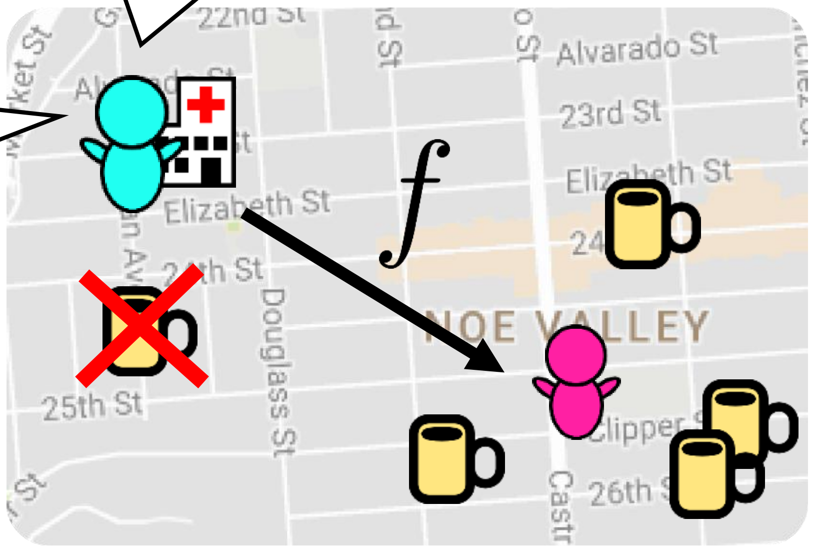
- Location information is sensitive.
- Solution: obfuscation mechanisms f

I want to use location services without disclosing my location



I'm at the fake location , closest ?

Here you go!

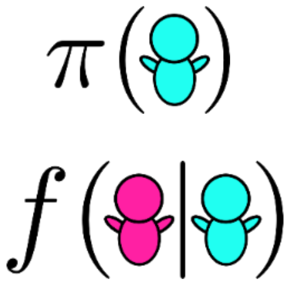


- We get some privacy.
- We lose some quality of service.
- There are many ways to evaluate the privacy and quality loss of obfuscation mechanisms. f

In this work

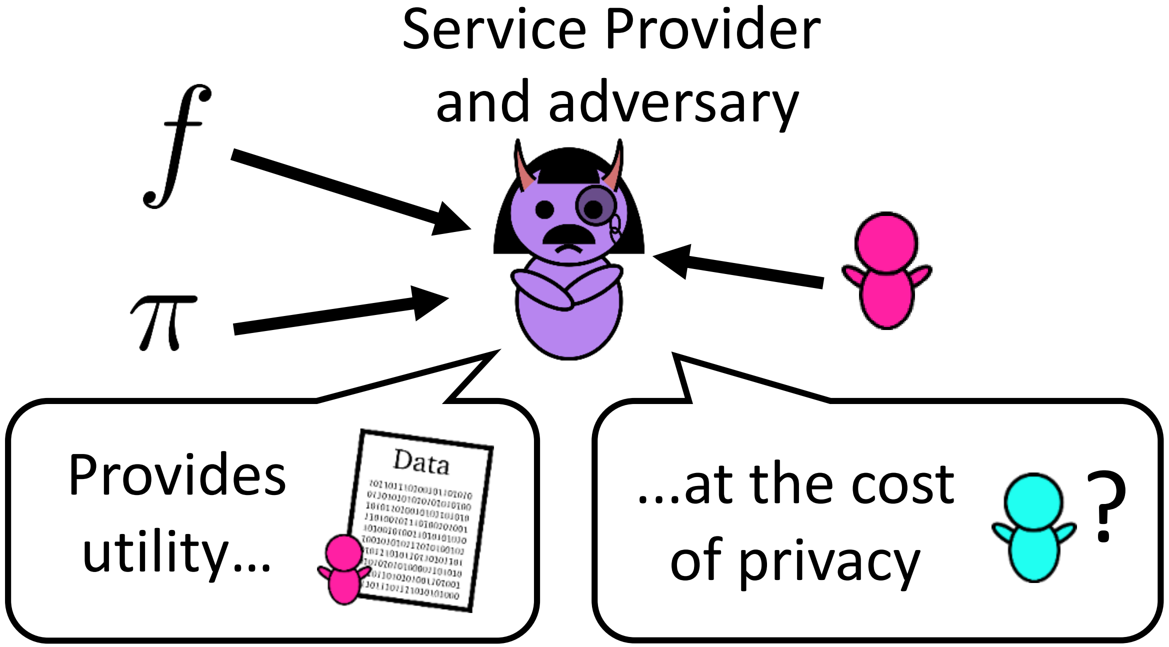
We study some flaws in the traditional evaluation approach and how to solve them.

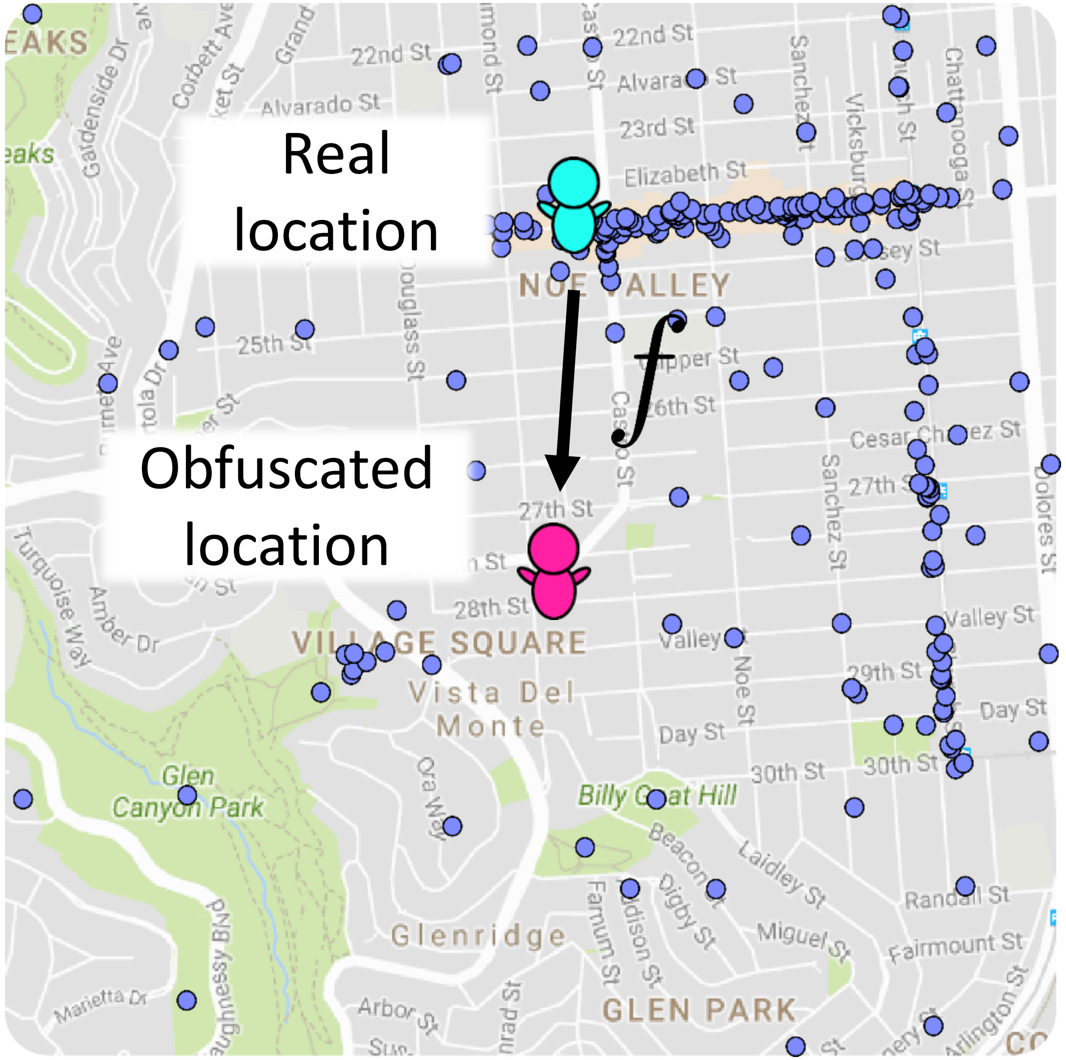
System Model



Prior of real locations

Obfuscation mechanism





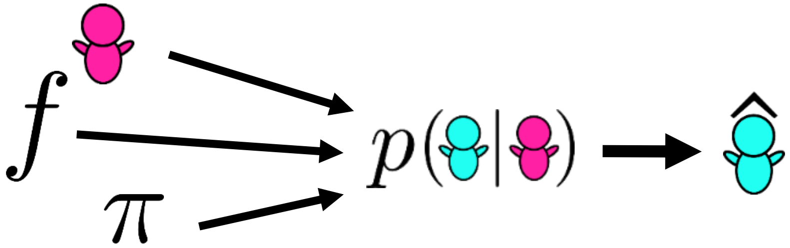
Traditional Evaluation: Metrics

- Quality Loss: **Average Loss**

$$\overline{\mathrm{Q}}(f, \pi) = \mathrm{E}\{d_Q(\cdot, \cdot)\}$$

Euclidean, Hamming, semantic, …

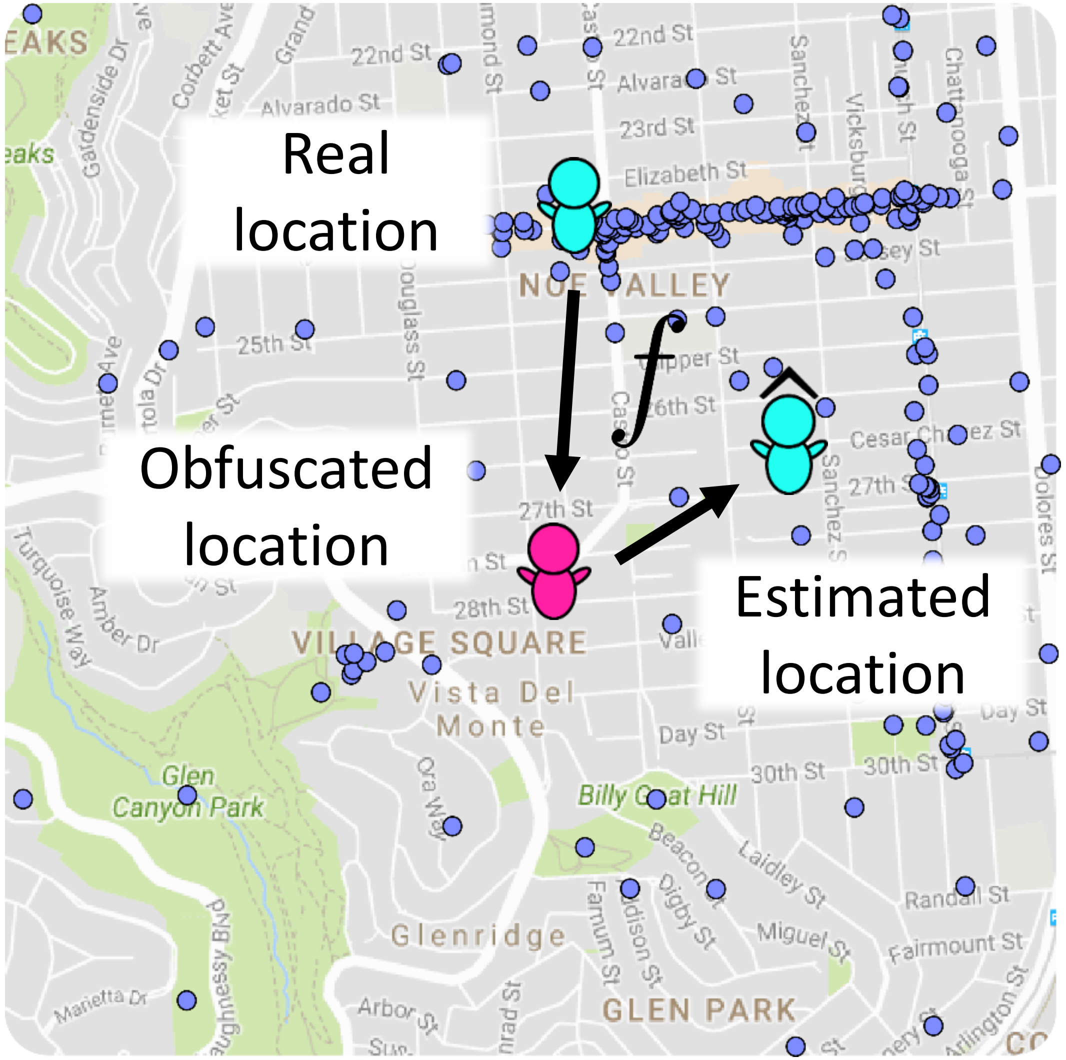
- Privacy: **Average Adversary Error**



Adversary’s estimation of the real location

$$\mathrm{P}_{\mathtt{AE}}(f, \pi) = \mathrm{E}\{d_P(\cdot, \hat{\cdot})\}$$

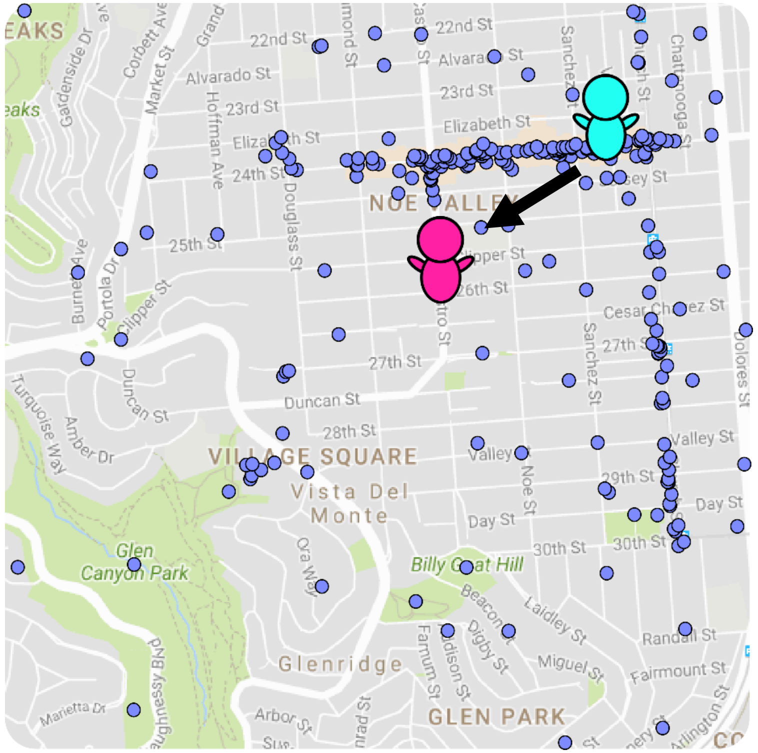
Euclidean, Hamming, semantic, …



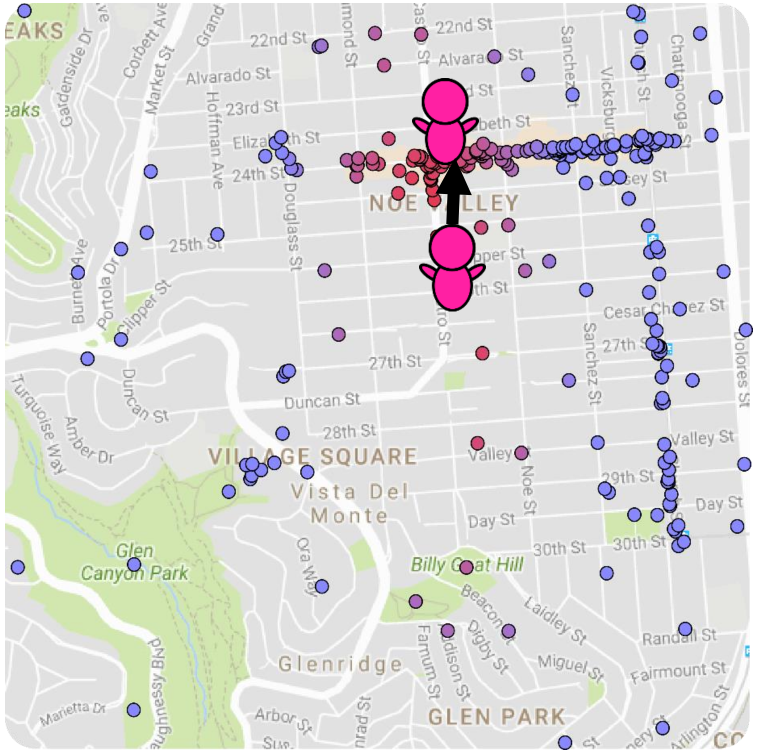
Shokri, Reza, et al. "Quantifying location privacy." *Security and privacy (sp), 2011 ieee symposium on*. IEEE, 2011.

Optimal Remapping [1]

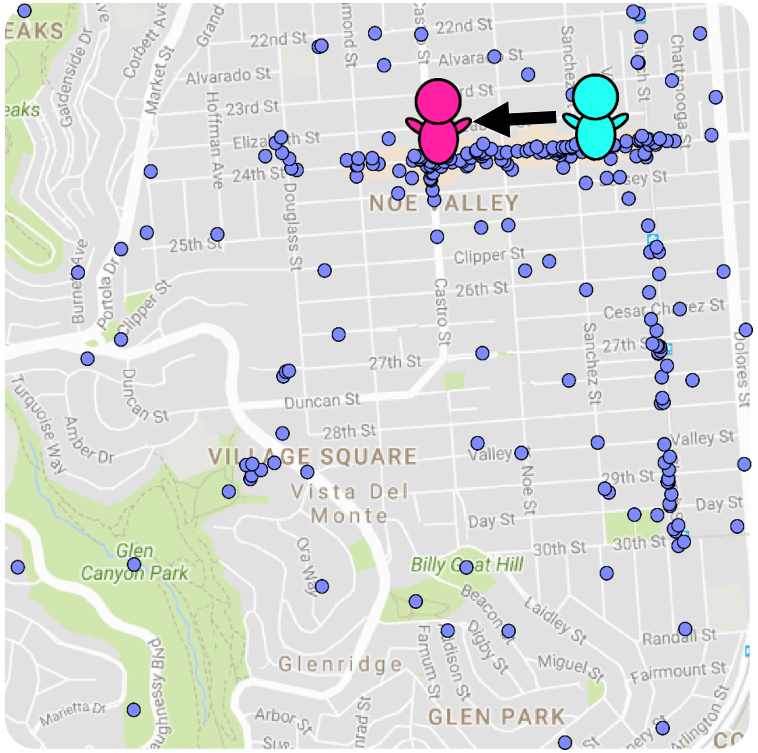
How to compute the optimal remapping of a mechanism f.



Step 1: Generate a random location using the mechanism



Step 2: Compute the posterior and remap to its “center”.



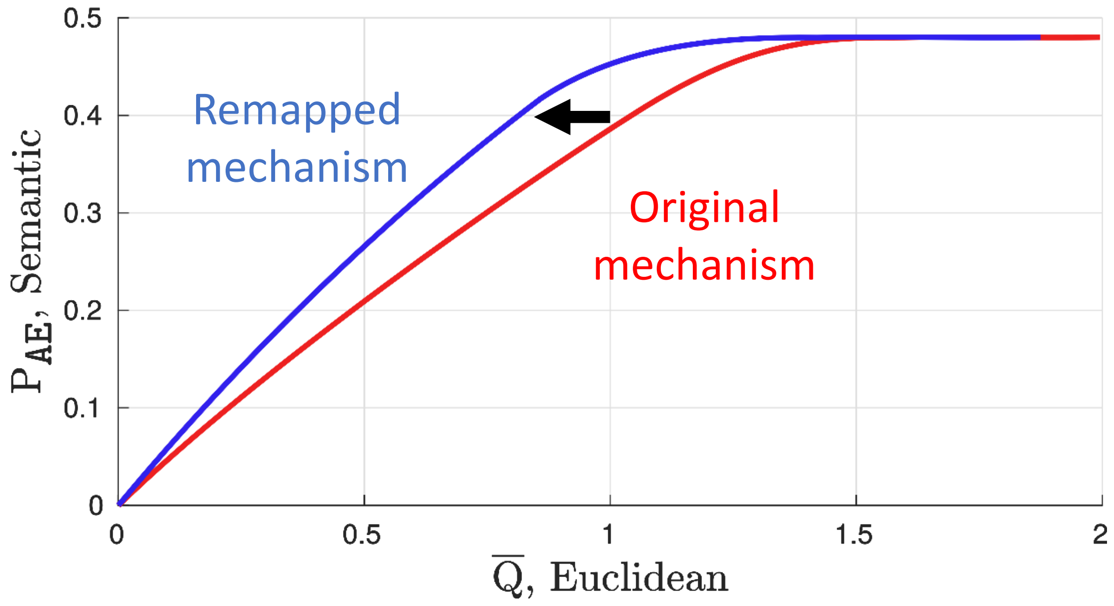
The generated output is the output after the remapping.

[1] Chatzikokolakis, K., Elsalamouny, E., & Palamidessi, C. “Efficient Utility Improvement for Location Privacy.” *PETS’17*.

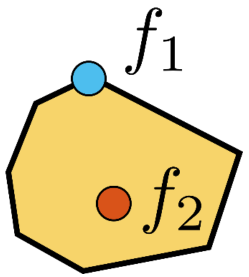
Traditional Evaluation: Example and Remapping

Traditional evaluation compares average error with average loss.

$$\overline{\mathrm{Q}}(f,\pi) = \mathrm{E}\{d_Q(\,,\,)\} \qquad \mathrm{P}_{\mathtt{AE}}(f,\pi) = \mathrm{E}\{d_P(\,,\,)\}$$



- **Theorem**: if $d_Q = d_P$, the optimal remapping gives an optimal mechanism in terms of $\mathrm{P}_{\mathtt{AE}}$ vs. $\overline{\mathrm{Q}}$.
- **Lemma**: the set of optimal mechanisms forms a convex polytope.

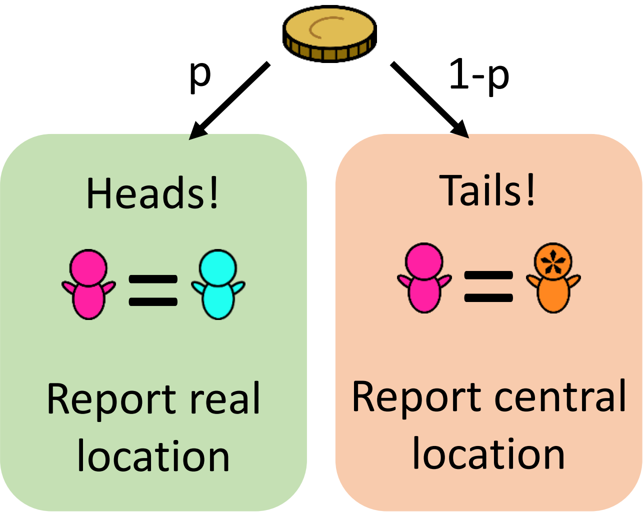


- This means there are many optimal mechanisms… are all of them “equally good”?

Problems of the Traditional Evaluation

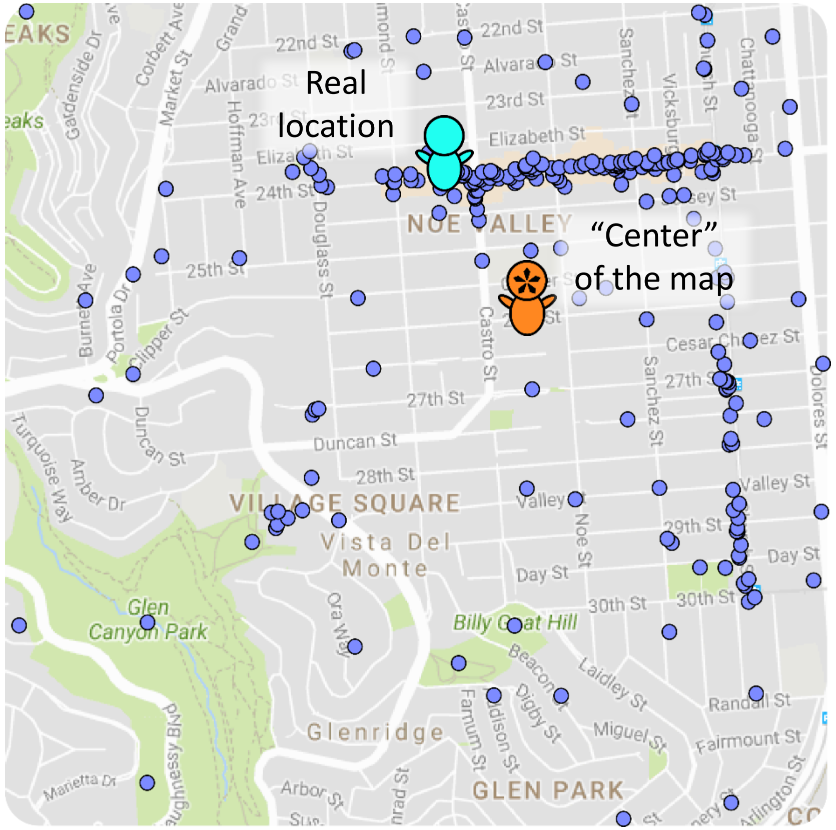
The Coin Mechanism

Flip a biased coin



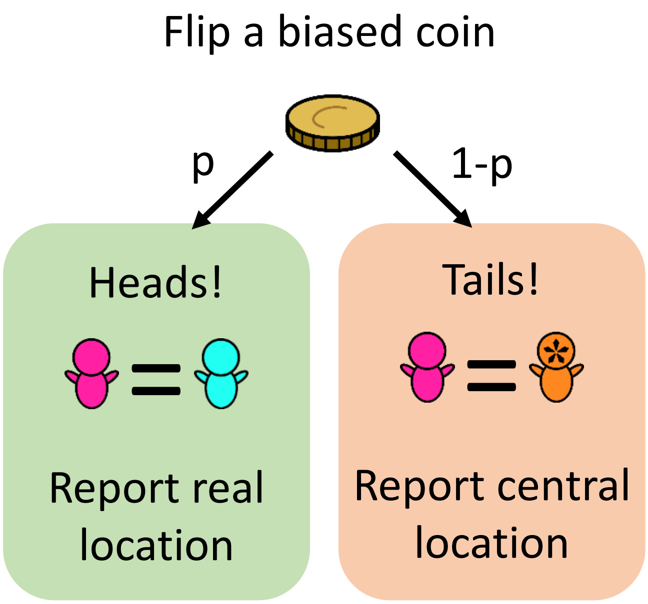
How “good” is this mechanism?

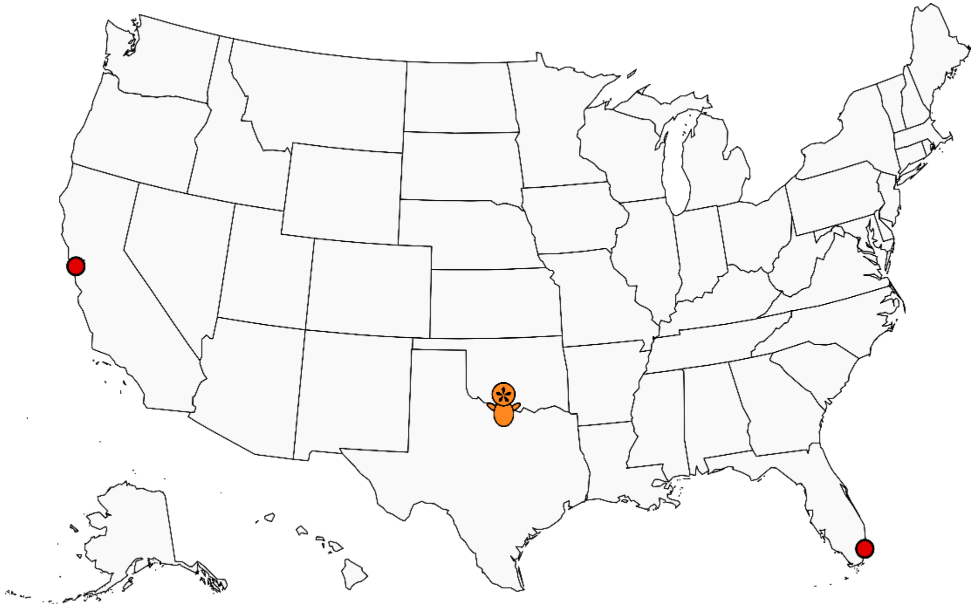
No privacy! Seems OK…



Problems of the Traditional Evaluation

The Coin Mechanism





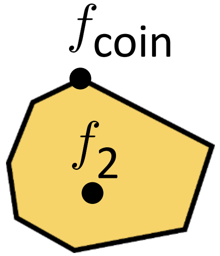
How “good” is this mechanism?

No privacy!

~~Seems OK…~~

No utility!

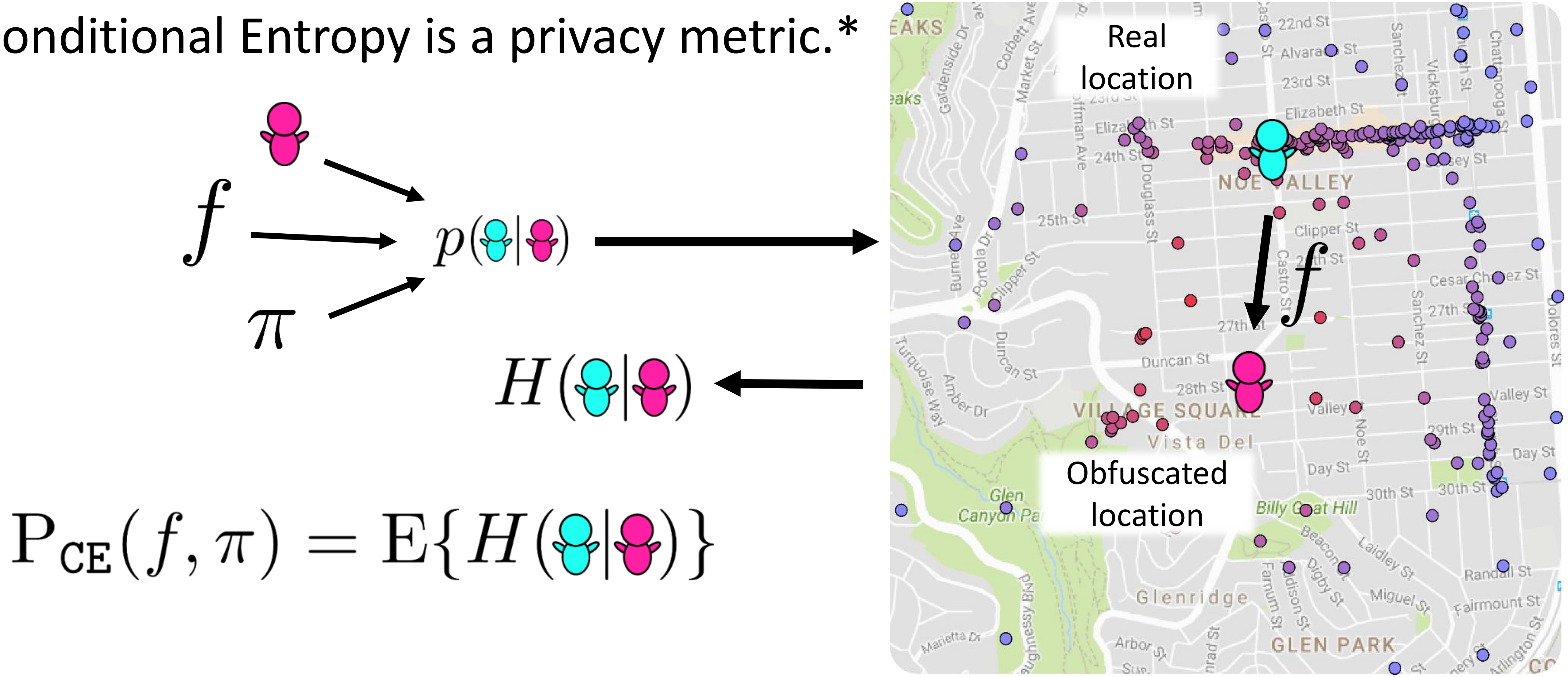
Polytope of optimal mechanisms



- The coin mechanism is useless in practice…
- … yet it is optimal in terms of P_{AE} vs. $\overline{\mathrm{Q}}$.
- How do we identify and avoid these “undesirable” mechanisms?
- Our proposal: use **additional** privacy and/or quality loss metrics.
- We will see two:
  - Conditional Entropy
  - Worst-Case Loss

Solution 1: Conditional Entropy

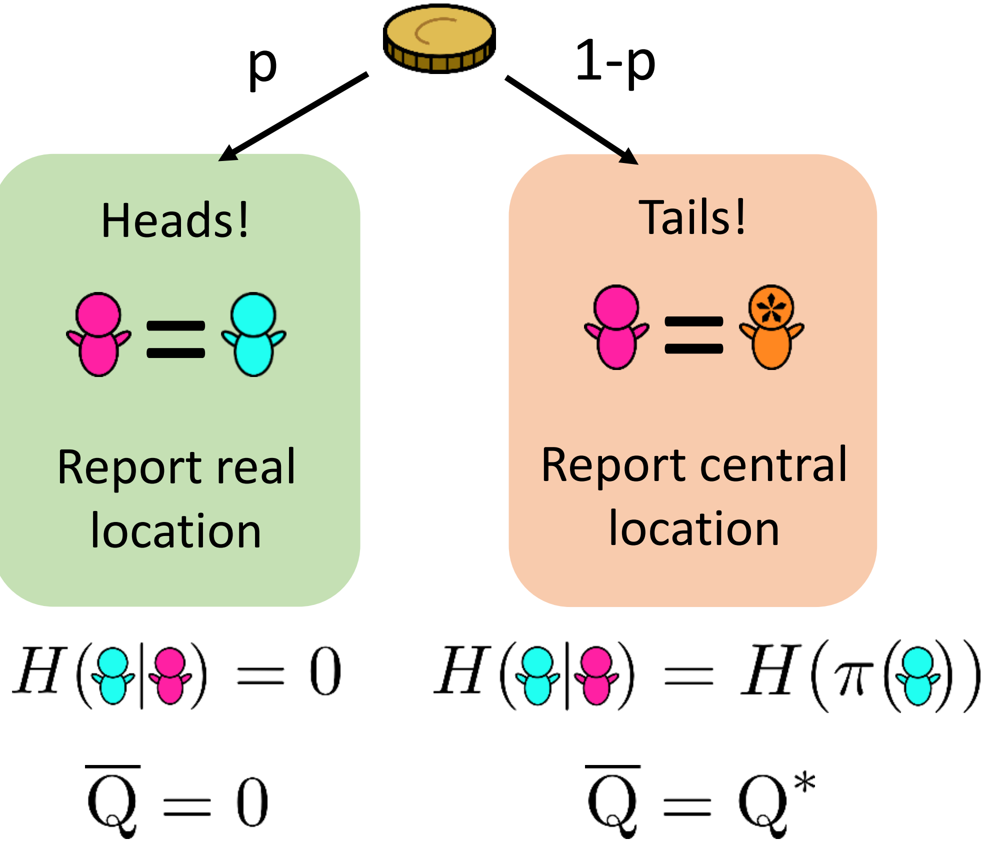
- The Conditional Entropy is a privacy metric.*

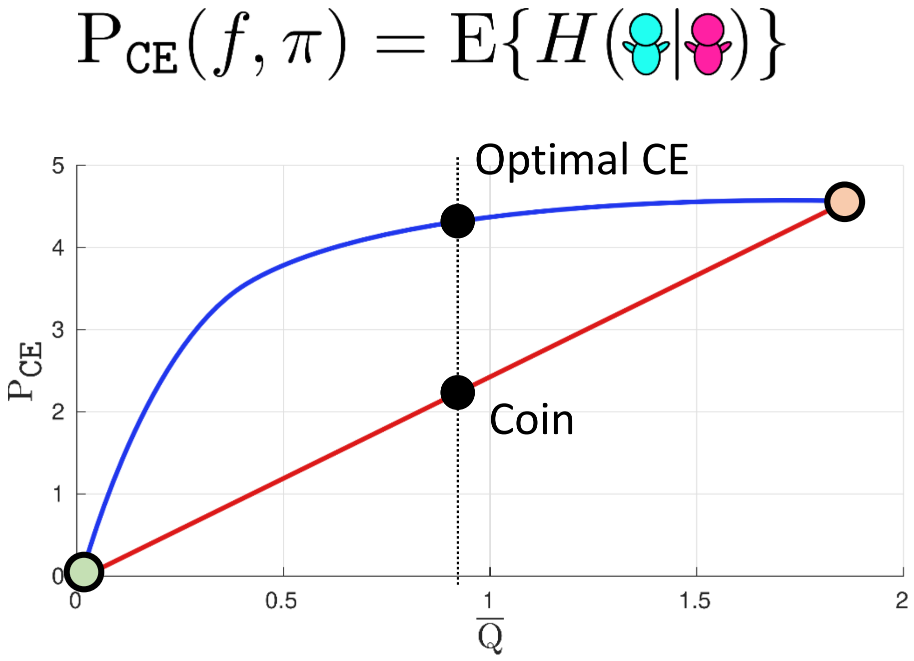


* Shokri, Reza, et al. "Quantifying location privacy." *Security and privacy (sp), 2011 ieee symposium on*. IEEE, 2011.

Conditional Entropy II

- How does it help us?

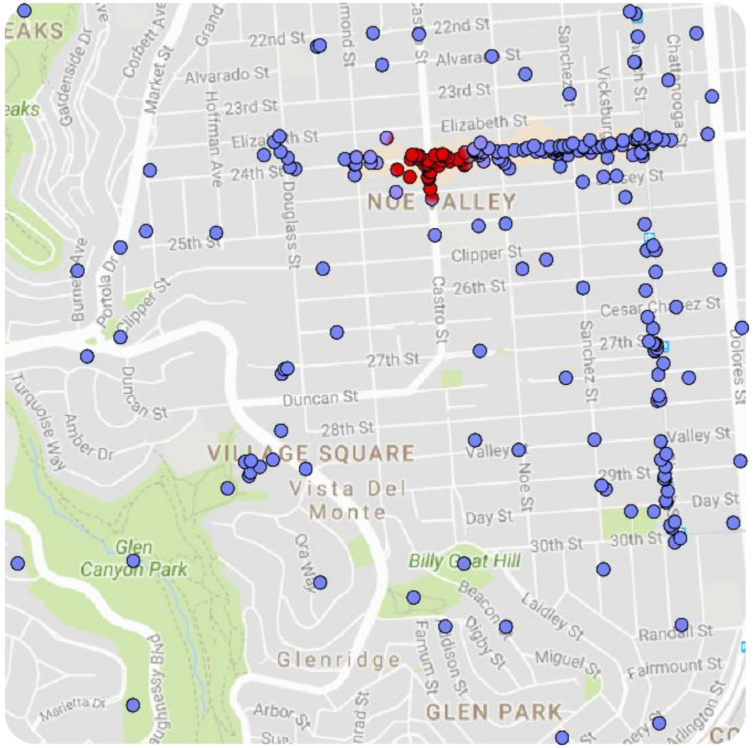




- The conditional entropy is concave!
- The coin performs poorly.
- The conditional entropy reveals "binary" mechanisms such as the coin.

Conditional Entropy III

- Is a mechanism that maximizes the conditional entropy “good” enough?
- Consider this adversary posterior:
- This is undesirable for the user… yet it achieves large conditional entropy.
- Therefore, we have to design mechanisms using CE as a **complementary metric**.



P_{CE} ↑ P_{AE} ↓

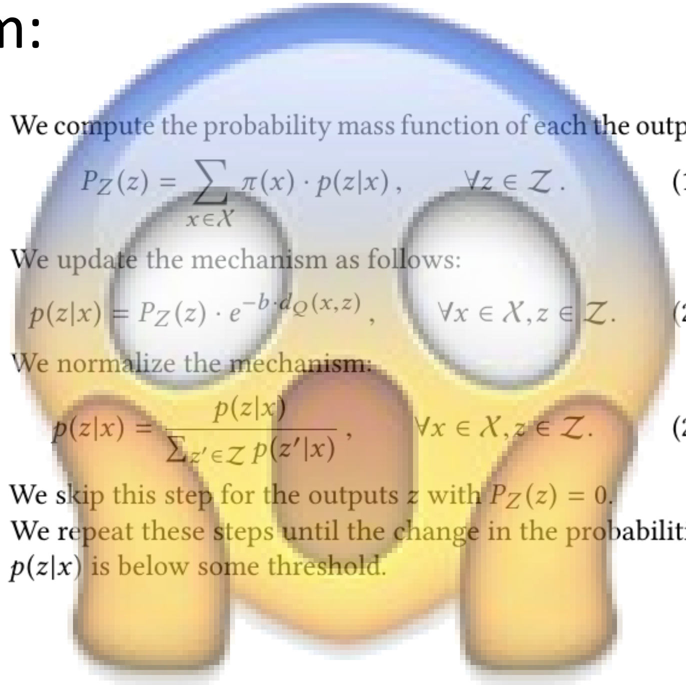
Conditional Entropy IV. Design.

- How to design a mechanism that performs well in terms of AE and CE?

$$\begin{aligned} \underset{f}{\text{maximize}} \quad & \mathrm{P}_{\mathtt{CE}}(f,\pi) \\ \text{s.t.} \quad & \overline{\mathrm{Q}}(f,\pi) \le \overline{\mathrm{Q}}_{\max} \\ & f \in \mathcal{P} \end{aligned} \longrightarrow \begin{aligned} \underset{f}{\text{minimize}} \quad & \mathrm{I}(\;;\;) \\ \text{s.t.} \quad & \overline{\mathrm{Q}}(f,\pi) \le \overline{\mathrm{Q}}_{\max} \\ & f \in \mathcal{P} \end{aligned} \longrightarrow$$

Rate-Distortion: Blahut-Arimoto

- Algorithm:



(1) We compute the probability mass function of each the output:

$$P_Z(z) = \sum_{x \in X} \pi(x) \cdot p(z|x)\,, \qquad \forall z \in \mathcal{Z}\,. \tag{19}$$

(2) We update the mechanism as follows:

$$p(z|x) = P_Z(z) \cdot e^{-b \cdot d_Q(x,z)}\,, \qquad \forall x \in \mathcal{X}, z \in \mathcal{Z}. \tag{20}$$

(3) We normalize the mechanism:

$$p(z|x) = \frac{p(z|x)}{\sum_{z' \in \mathcal{Z}} p(z'|x)}\,, \qquad \forall x \in \mathcal{X}, z \in \mathcal{Z}. \tag{21}$$

We skip this step for the outputs z with $P_Z(z) = 0$.

(4) We repeat these steps until the change in the probabilities $p(z|x)$ is below some threshold.

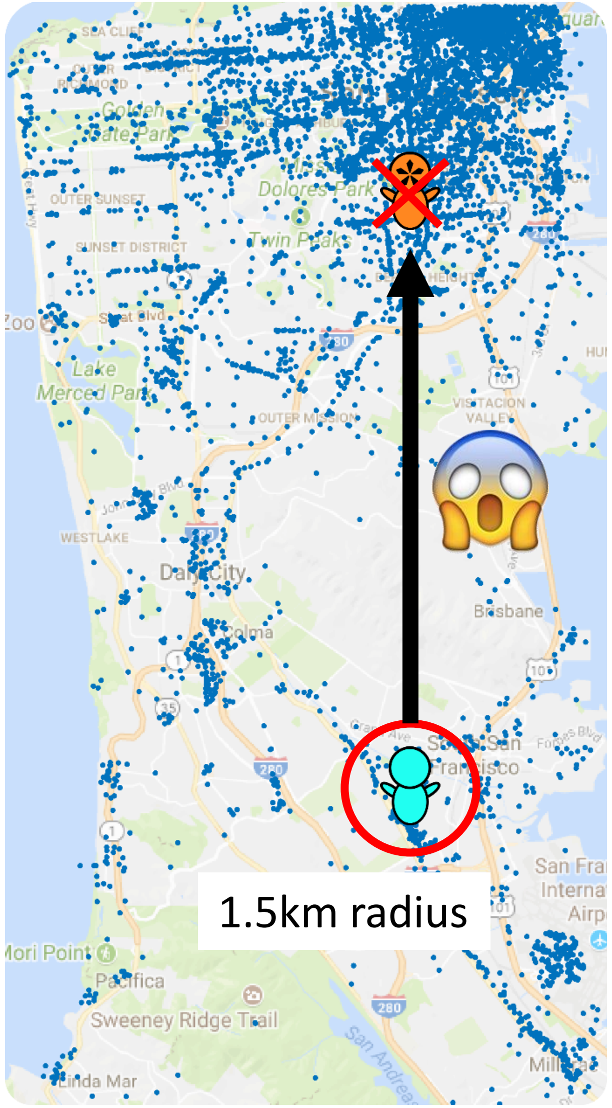
Summary:

- Tries to make an exponential posterior (we call it **ExPost**).
- For computational reasons, we need to perform approximations.
- The more computational power we have, the closer it is to the optimal mechanism in terms of CE.
- Iterative.
- Uses remapping to achieve optimal AE.

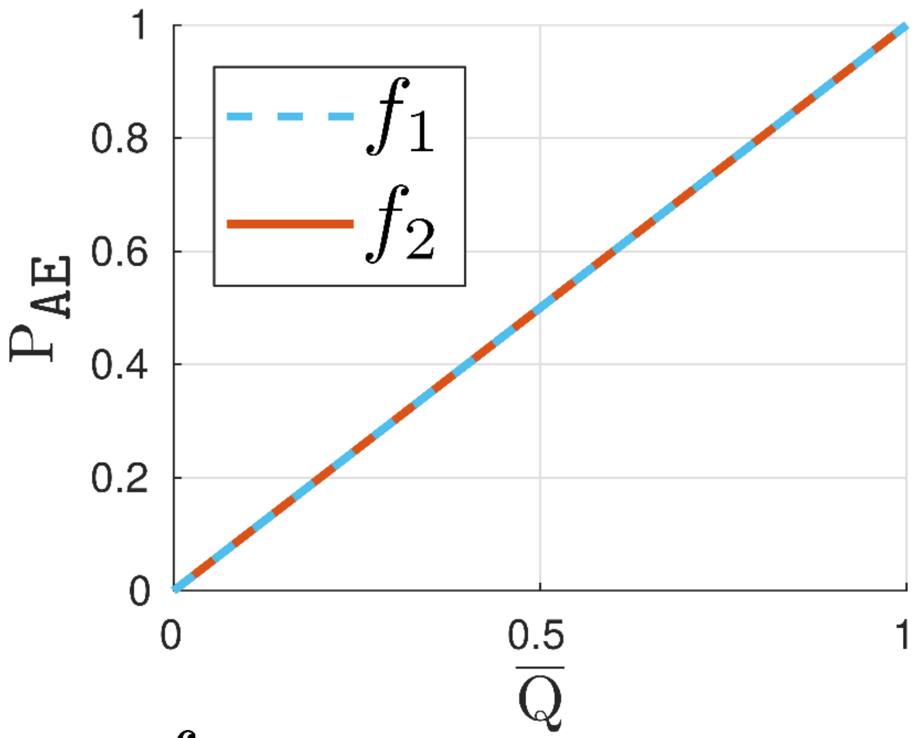
Solution 2: Worst-Case Loss

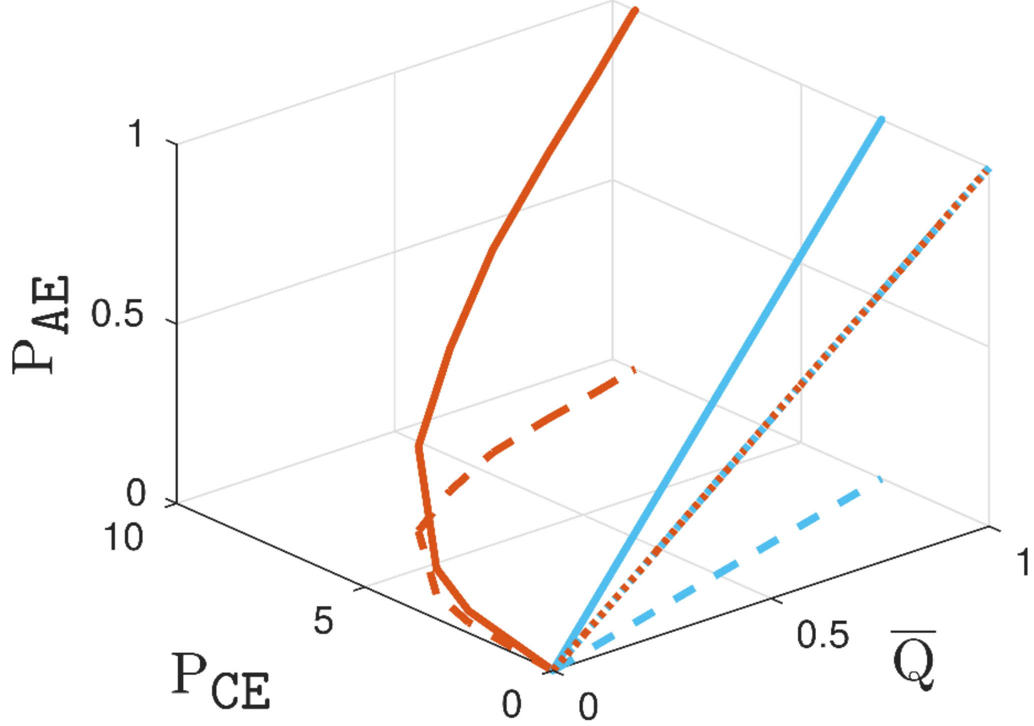
$$Q^{+}(f,\pi) = \max_{\substack{\pi(\cdot)>0 \\ f(\cdot|\cdot)>0}} d_Q(\cdot,\cdot)$$

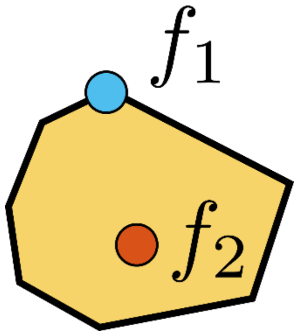
- How does it help us?
- Tails → Huge loss
- Having a constraint on the WC loss avoids this.
- This constraint makes sense in real applications where we need a minimum utility (e.g., search nearby points of interest).
- Implementation: add a WC loss constraint to the design problem, use truncation, etc.



Multi-Dimensional Notion of Privacy







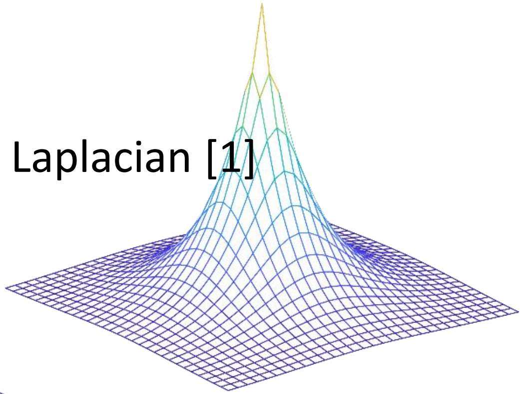
- Both mechanisms are optimal with respect to this privacy and quality loss notions.

- The two-dimensional approach is misleading.
- Consider privacy as a **multi-dimensional notion**.

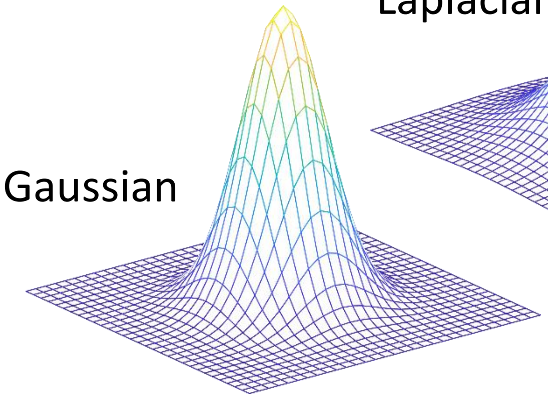
Evaluation I. Mechanisms.

- Selection of relevant mechanisms.

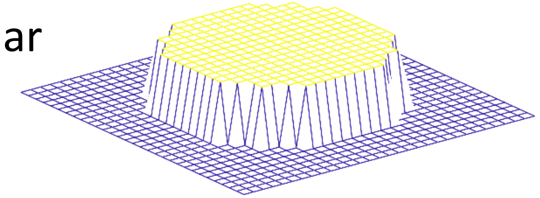
Laplacian [1]

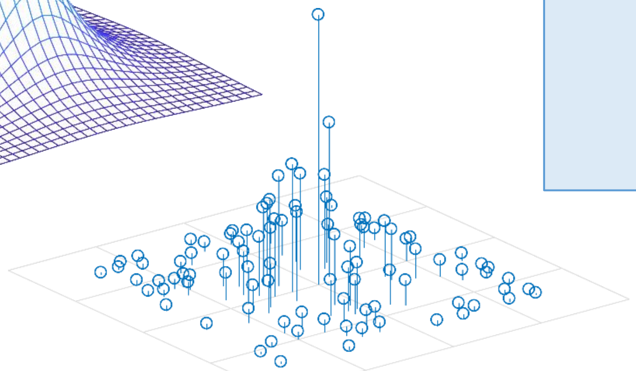


Gaussian



Circular





Exponential

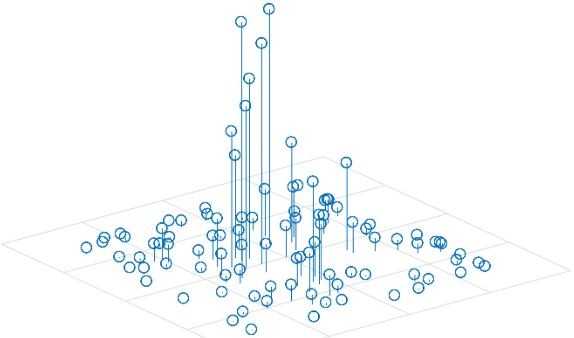
We also perform an optimal remapping after these mechanisms to improve them.

Optimal AE [2]

$$\begin{aligned} \underset{f}{\text{maximize}} \quad & \mathrm{P}_{\mathtt{AE}}(f,\pi) \\ \text{s.t.} \quad & \overline{\mathrm{Q}}(f,\pi) \leq \overline{\mathrm{Q}}_{\max} \\ & f \in \mathcal{P} \end{aligned}$$

Linear program!
Only feasible in simple scenarios.

- Two from our work



Exponential Posterior (ExPost)

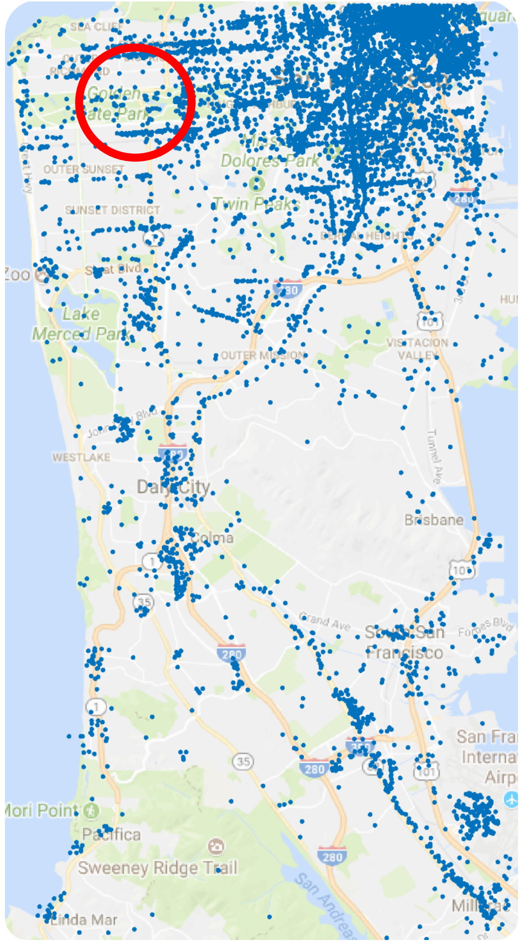
The coin

[1] Chatzikokolakis, K., Elsalamouny, E., & Palamidessi, C. “Efficient Utility Improvement for Location Privacy.” *PETS’17*.

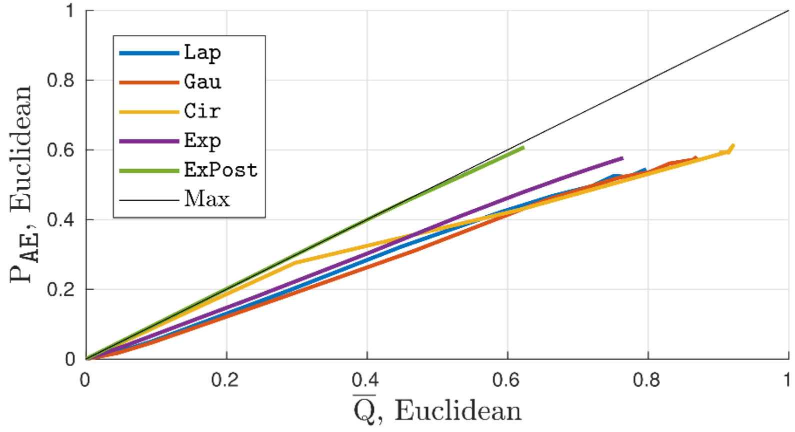
[2] Shokri, Reza, et al. "Protecting location privacy: optimal strategy against localization attacks." *CCS’12*

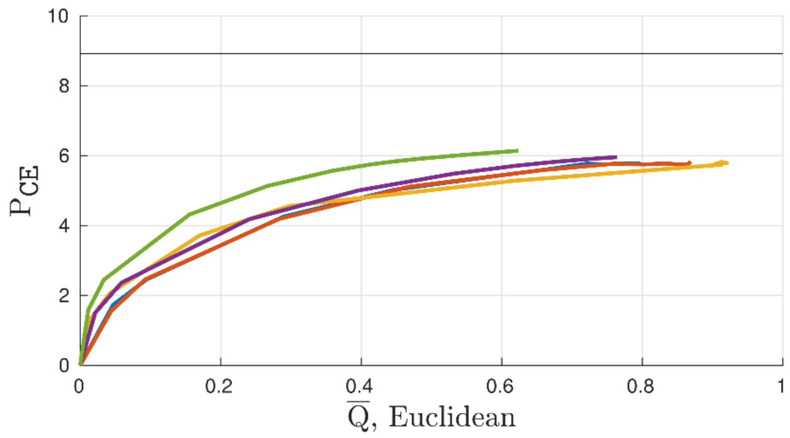
Evaluation II. Continuous Scenario.

Datasets: Gowalla, Brightkite
San Francisco region

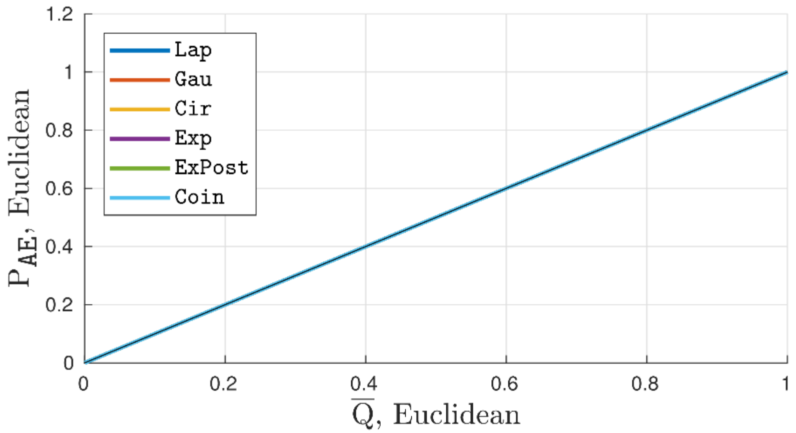


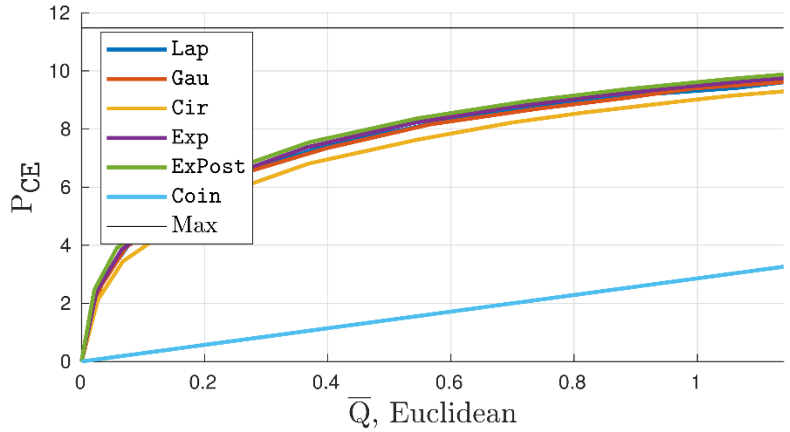
With Worst-Case Loss = 1.5km





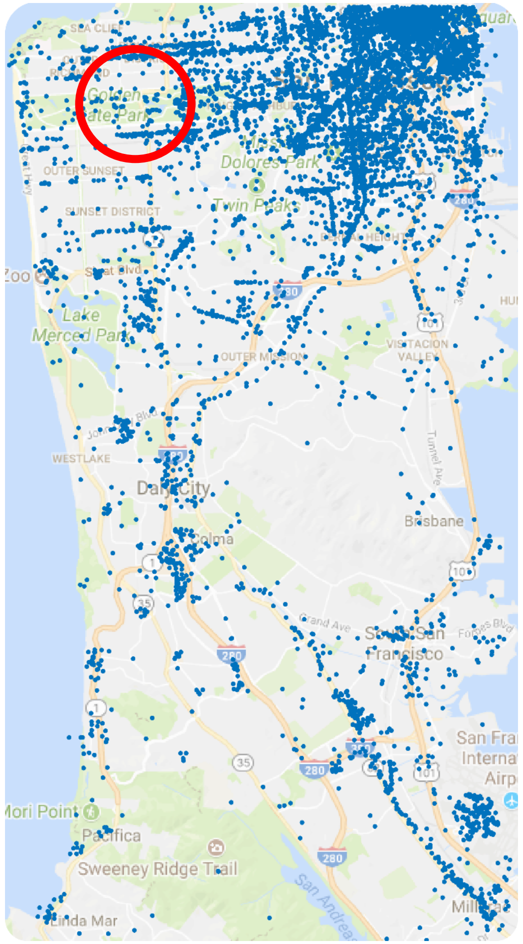
Without Worst-Case Loss





Evaluation II. Continuous Scenario.

Datasets: Gowalla, Brightkite
San Francisco region

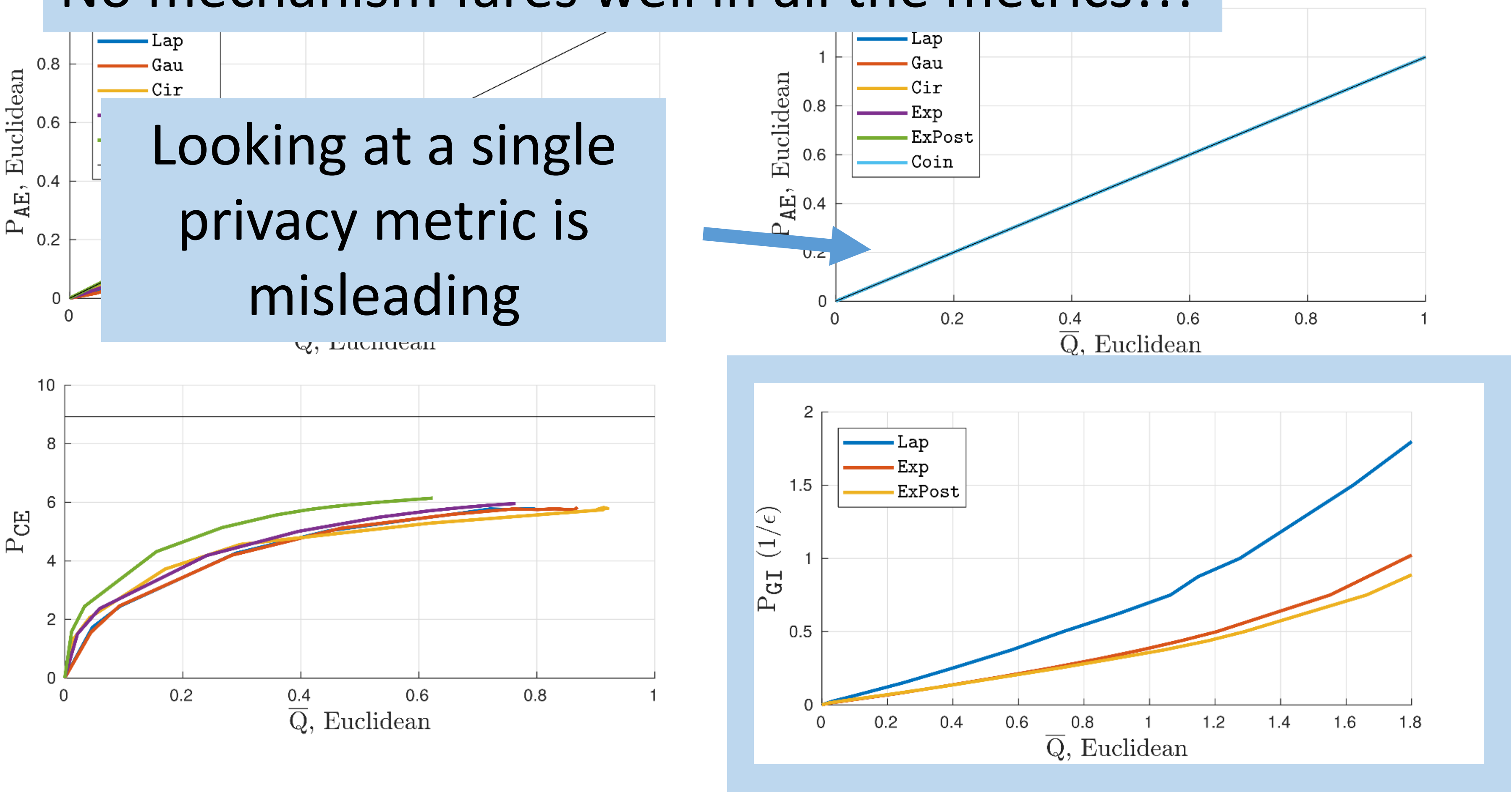


With Worst-Case Loss = 1.5km

Without Worst-Case Loss

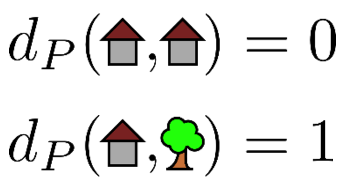
No mechanism fares well in all the metrics!!!

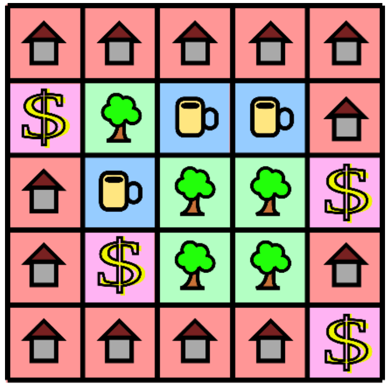
Looking at a single privacy metric is misleading



Evaluation III. Discrete Scenario (Semantic)

- We consider a semantic metric.

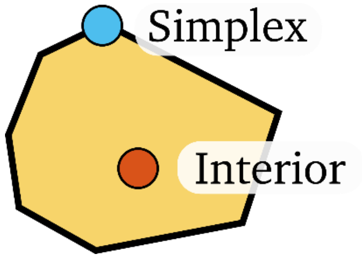


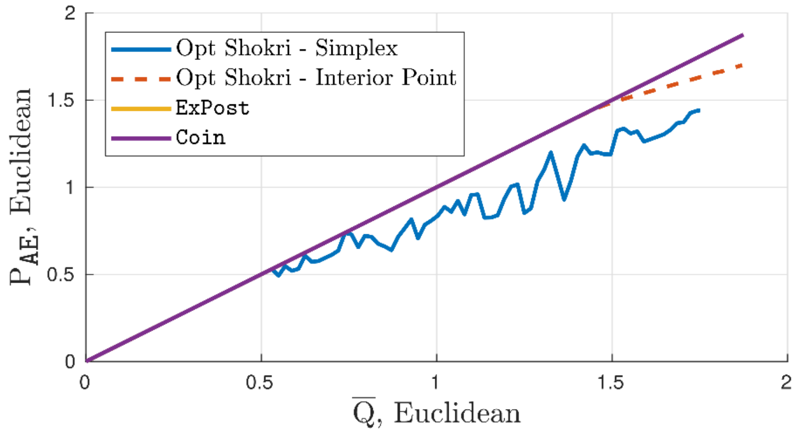


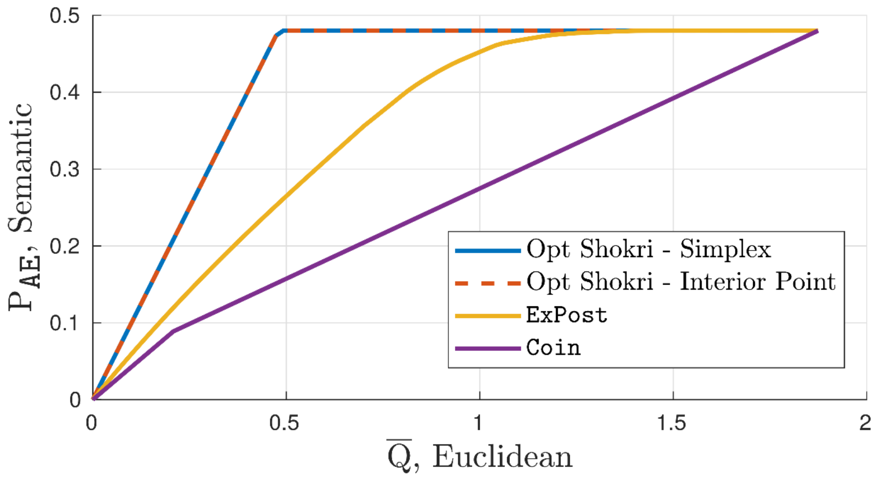
Home
Park
Shop
Café

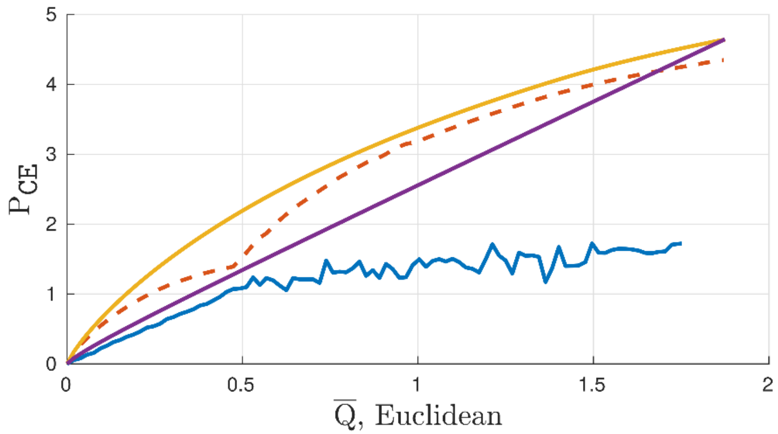
- We evaluate *Shokri et. al* optimal mechanism [2], optimized for the semantic metric.

$$
\begin{aligned}
\underset{f}{\text{maximize}}\quad & \mathrm{P}_{\mathtt{AE}}(f,\pi) \\
\text{s.t.}\quad & \overline{\mathrm{Q}}(f,\pi) \le \overline{\mathrm{Q}}_{\max} \\
& f \in \mathcal{P}
\end{aligned}
$$





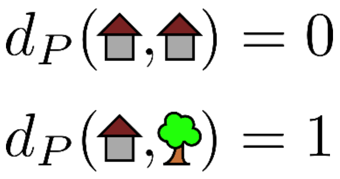


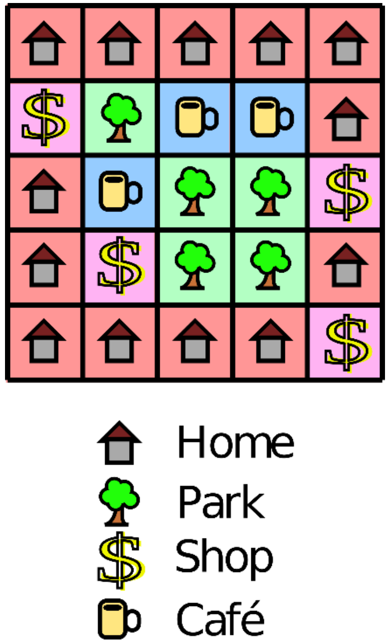


[2] Shokri, Reza, et al. "Protecting location privacy: optimal strategy against localization attacks." *CCS'12*

Evaluation III. Discrete Scenario (Semantic)

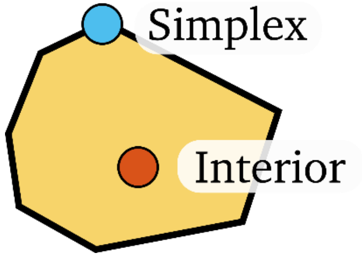
- We consider a semantic metric.

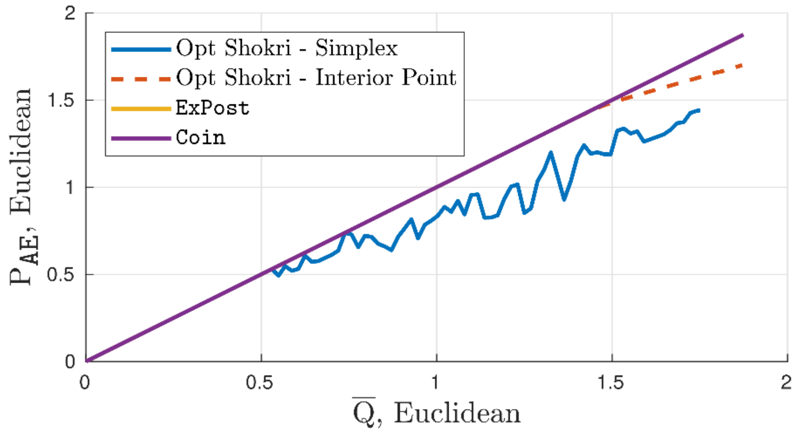




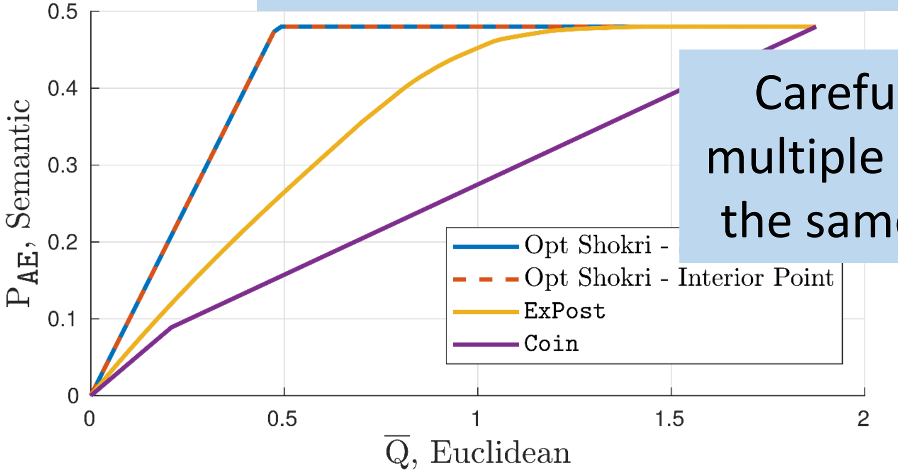
- We evaluate *Shokri et. al* optimal mechanism [2], optimized for the semantic metric.

$$\begin{aligned} \underset{f}{\text{maximize}}\quad & \mathrm{P}_{\mathtt{AE}}(f,\pi) \\ \text{s.t.}\quad & \overline{\mathrm{Q}}(f,\pi) \le \overline{\mathrm{Q}}_{\max} \\ & f \in \mathcal{P} \end{aligned}$$

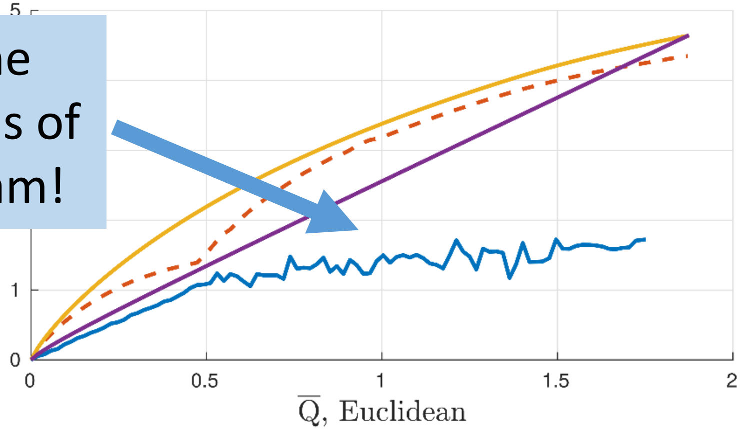




No mechanism fares well in all the metrics!!!



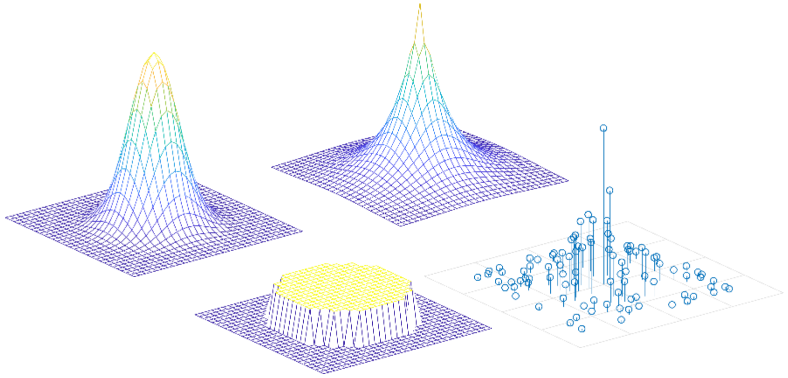
Careful with the multiple solutions of the same program!



[2] Shokri, Reza, et al. "Protecting location privacy: optimal strategy against localization attacks." *CCS'12*

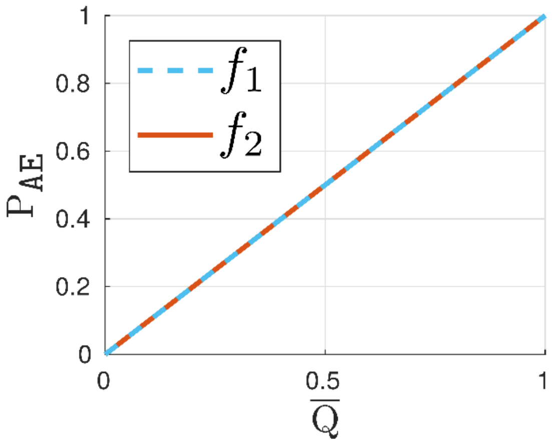
Conclusions

Many location-privacy mechanisms are being proposed

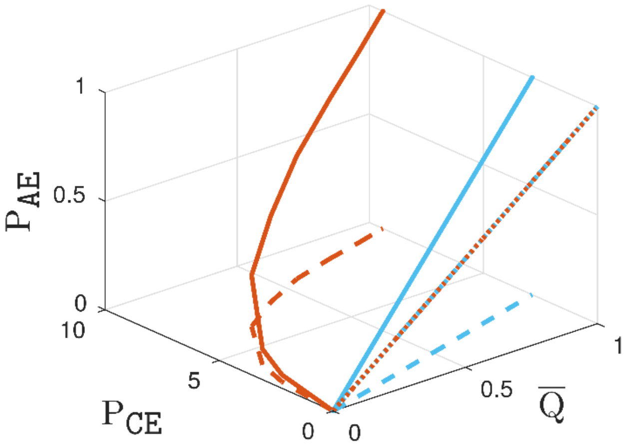


$$\begin{aligned} &\underset{f}{\text{maximize}} \quad \mathrm{P}_{\mathrm{AE}}(f,\pi) \\ &\text{s.t.} \quad \overline{\mathrm{Q}}(f,\pi) \leq \overline{\mathrm{Q}}_{\max} \\ &\quad\quad\ f \in \mathcal{P} \end{aligned}$$

Most of them are evaluated following a two-dimensional approach



This might give “bad” mechanisms. Design and evaluation should be done considering privacy as a **multidimensional notion**.



Thank you!!

simonoya@gts.uvigo.es