

AtlantTIC

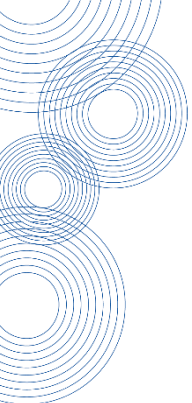
Research Center for
Information & Communication Technologies

Location Privacy. Where do we stand and where are we going?

Fernando Pérez-González

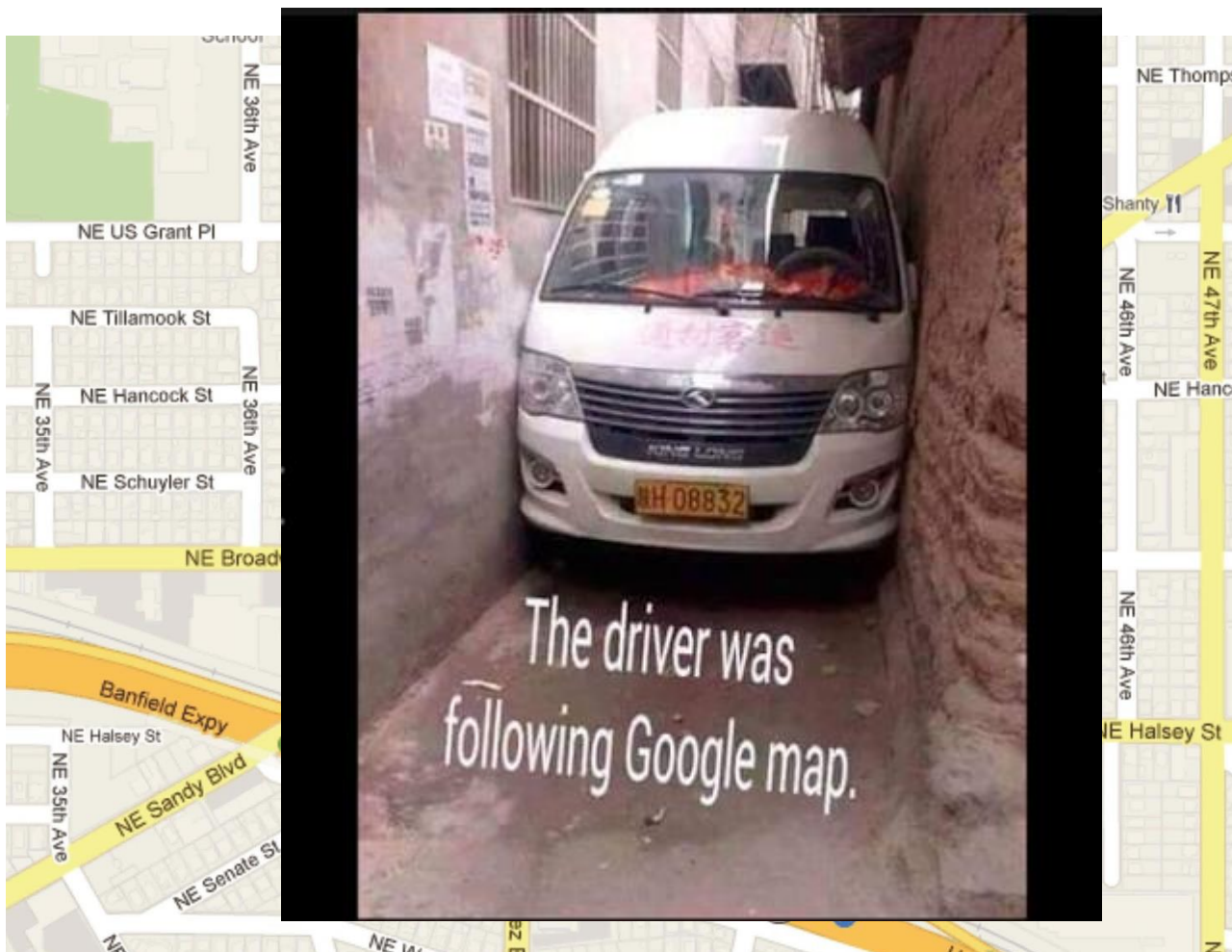
***Signal Theory and
Communications
Department***

***Universidad de Vigo -
SPAIN***

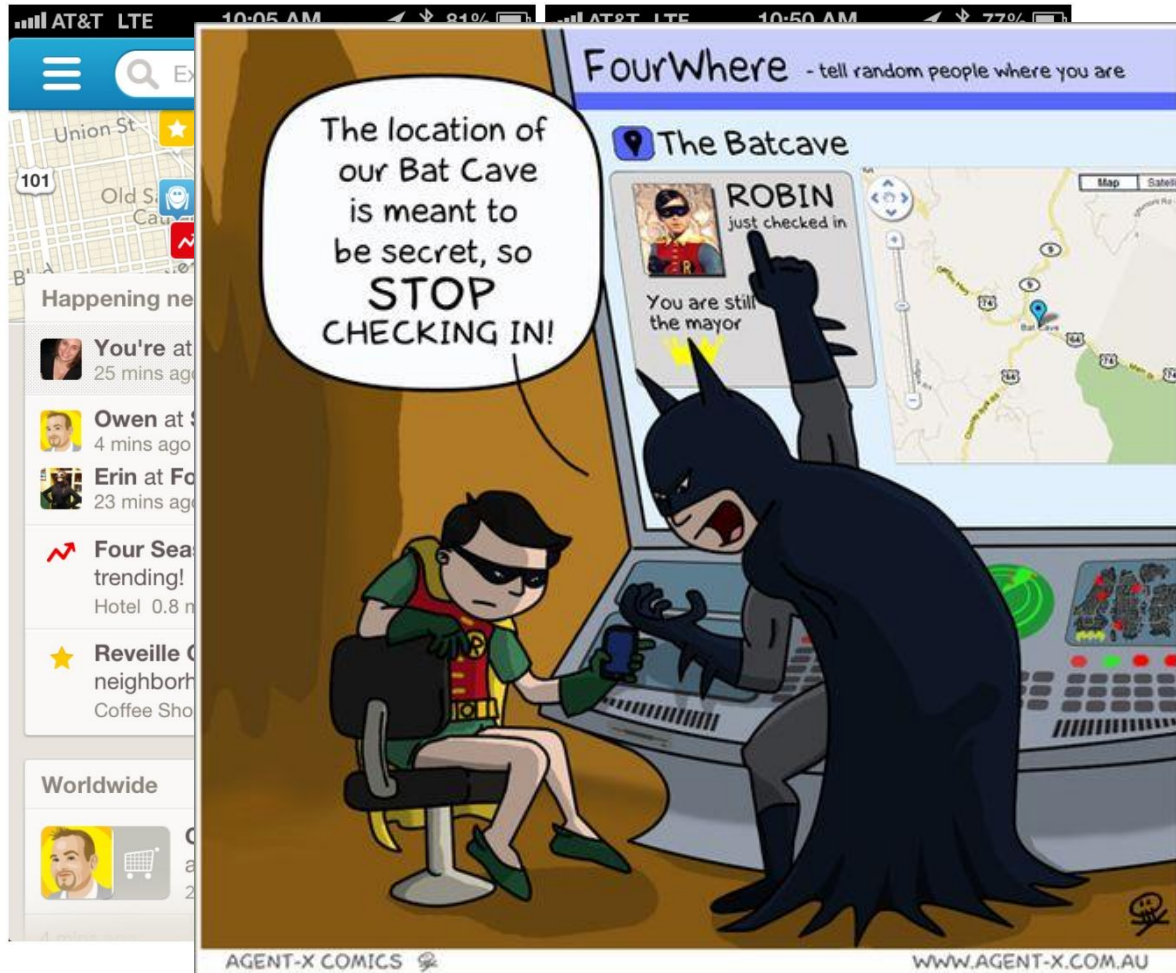


Why do we like location based apps?

Google maps



Foursquare



Facebook place tips

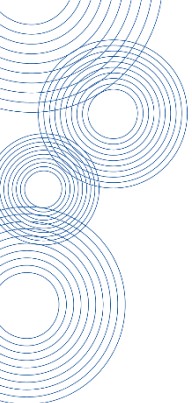


Waze



And, of course...





How can you be geolocated? (without you fully knowing)

IP-based Geolocation

The screenshot displays the GeoIP Tool interface. On the left, a search bar labeled 'Host/IP' contains the IP address 109.73.65.211. Below the search bar, the following geolocation data is listed:

- Nombre Host: 109.73.65.211
- Dirección de IP: 109.73.65.211
- País: United Kingdom
- Código de país: GB (GBR)
- Region: London 1068,GB,I1,"Luton
- Ciudad: London
- Hora local: 31 Oct 10:23 (GMT+0000)
- Código Postal: EC4N
- Latitud: 51.5144
- Longitud: -0.0941

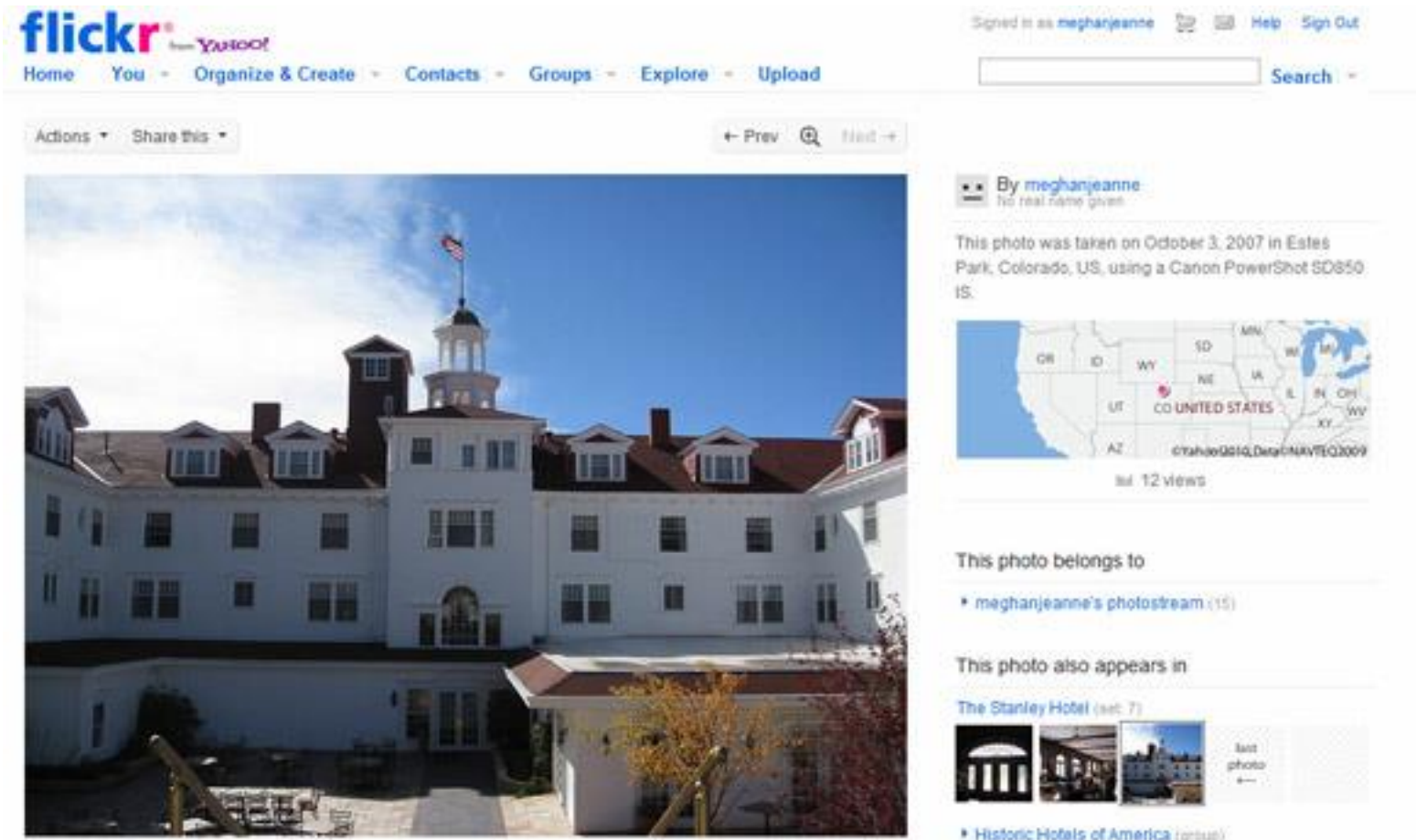
On the right, a Google Map of Europe shows a red pin in London. A tooltip above the pin displays the following information:

- País: United Kingdom
- Ciudad: London
- Dirección de IP: 109.73.65.211

The map also shows various European countries and cities, and includes a scale bar for 500 km.

Source: GeoIPTool

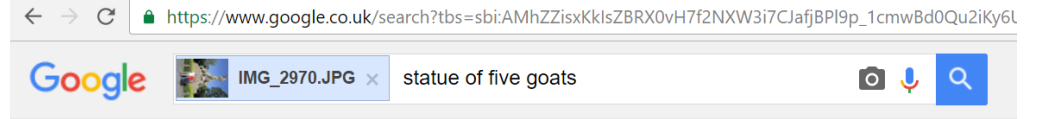
Meta-data based Geolocation



The screenshot shows a Flickr page for a photo of a large, white, multi-story building with a central tower and a cupola. The page includes the Flickr navigation bar, a search bar, and a metadata section. The metadata section contains the following information:

- By meghanjeanne** (no real name given)
- This photo was taken on October 3, 2007 in Estes Park, Colorado, US, using a Canon PowerShot SD850 IS.
- A map of the United States with a red pin in Colorado, showing 12 views.
- This photo belongs to [meghanjeanne's photostream](#) (15)
- This photo also appears in [The Stanley Hotel](#) (set: 7)
- [Historic Hotels of America](#) (group)

Landmark recognition Geolocation



All **Images** Maps Shopping More Search tools

About 2 results (0.56 seconds)



Image size:
4320 x 3240

No other sizes of this image found.

Best guess for this image: **statue of five goats**

[Legend of 5 goats | - Guangzhou.chn.info](#)

www.guangzhou.chn.info/overview/legend-of-5goats.html

There are many goat statues in Guangzhou and the **Statue of the Five Goats** is the most impressive, and now the one which were built in Yuexiou Park in 1959 ...

[Five Rams Statue, Guangzhou 24 Insider Tips, Photos and Reviews](#)

https://www.virtualltourist.com/.../Things_To_Do-Guangzhou-Five_Rams_Statue-BR-...

Five Rams Statue reviews and photos from real travelers and locals in Guangzhou, ... The **statue of five goats** was originally created by Guangzhou sculptors Yin ...

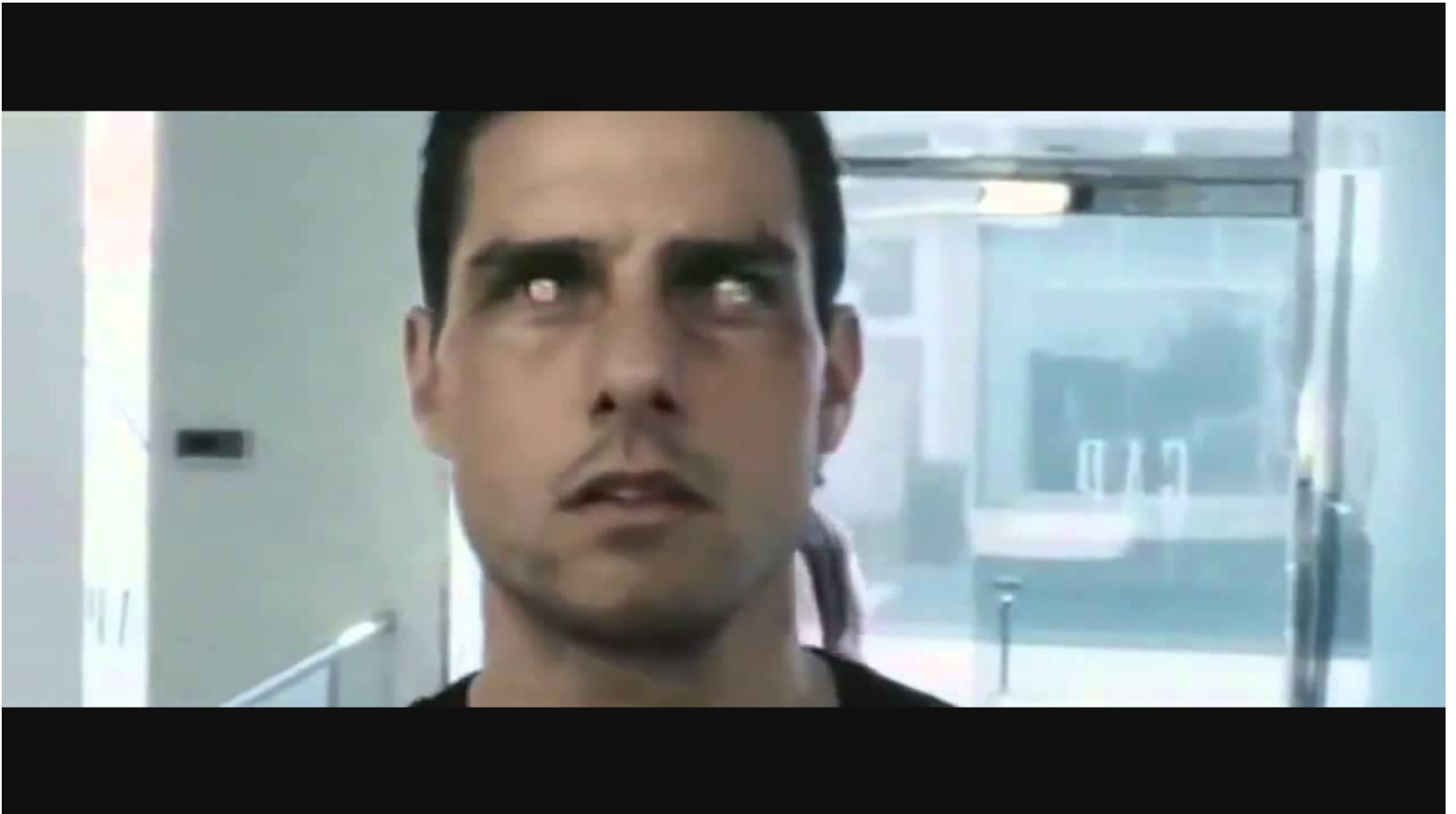
Visually similar images

Report images





Biometric geolocation



Credit card usage Geolocation



MailOnline



FREE - On the Microsoft Store

Mastercard under fire for tracking customer credit card purchases to sell to advertisers

- Credit card firm refuses to reveal 'proprietary' technique that allows it to anonymously track customers and target them with online ads
- Privacy campaigners accuse firm of 'treating details of our personal behaviour like their own property'
- System tracks information about the date, time, amount and merchant
- Credit card firm says system is only operational in US

By MARK PRIGG

PUBLISHED: 15:52, 17 October 2012 | UPDATED: 17:36, 17 October 2012



View comments

Mastercard has come under fire for tracking its US customer's purchases and selling the data to advertisers.

The credit card company's MasterCard Advisors Media Solutions Group boasts it can target the most affluent customers and tell advertisers who is most likely to buy their products.

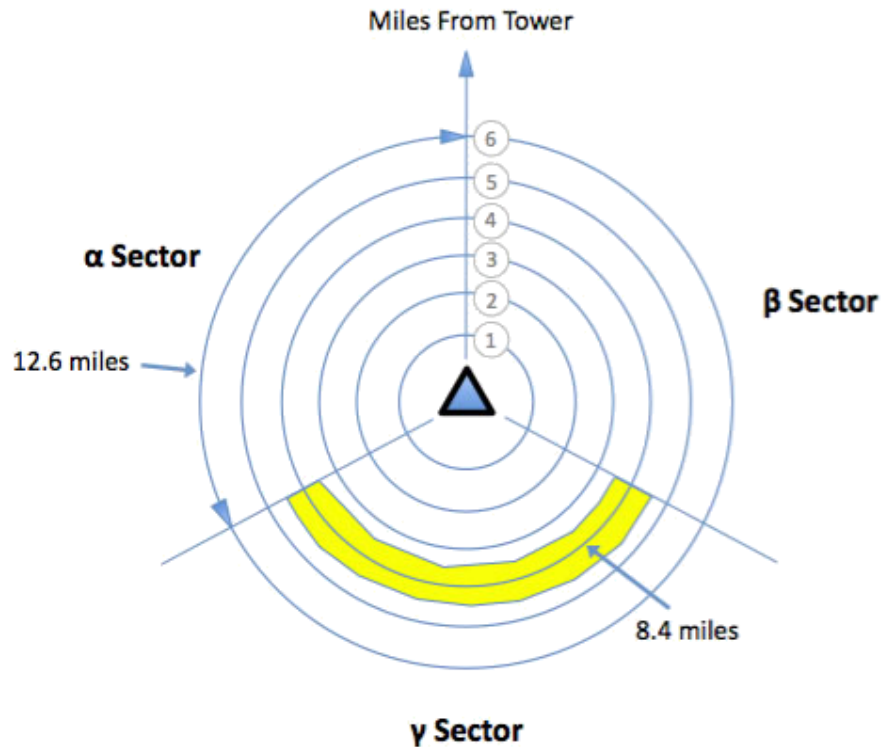
The firm does this by tracking a consumer's credit card details - although it says their identity



Triangulation and other geolocation techniques

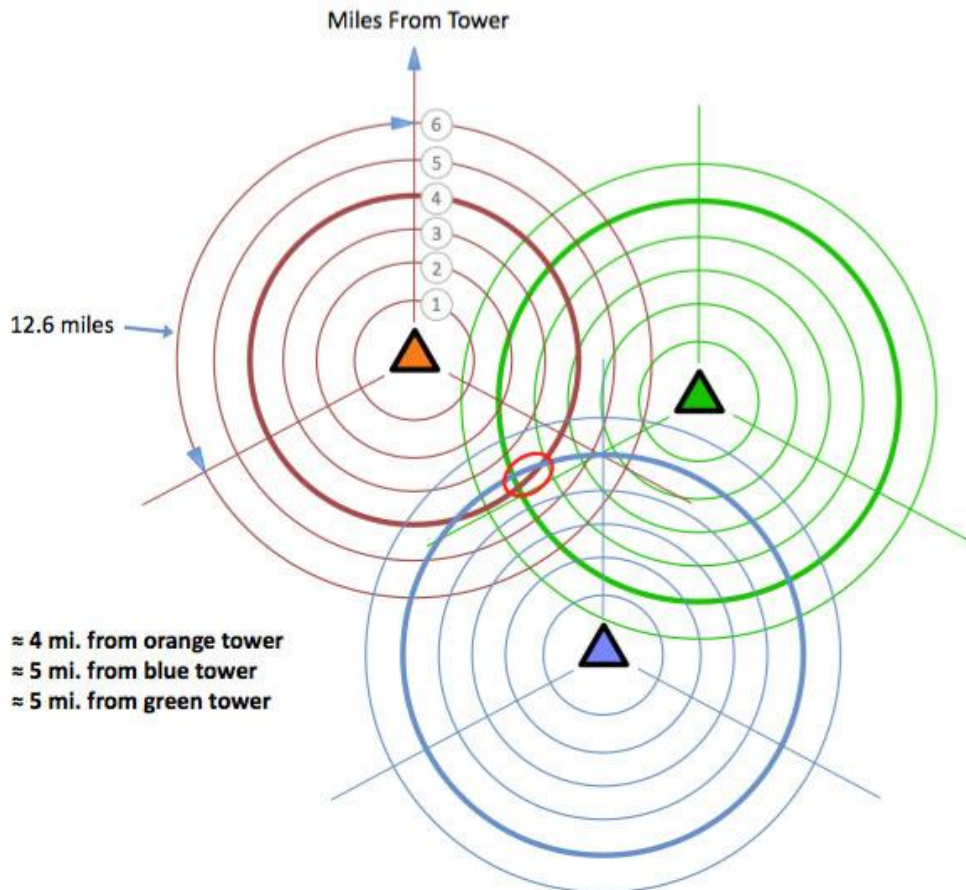


Signal strength-based triangulation



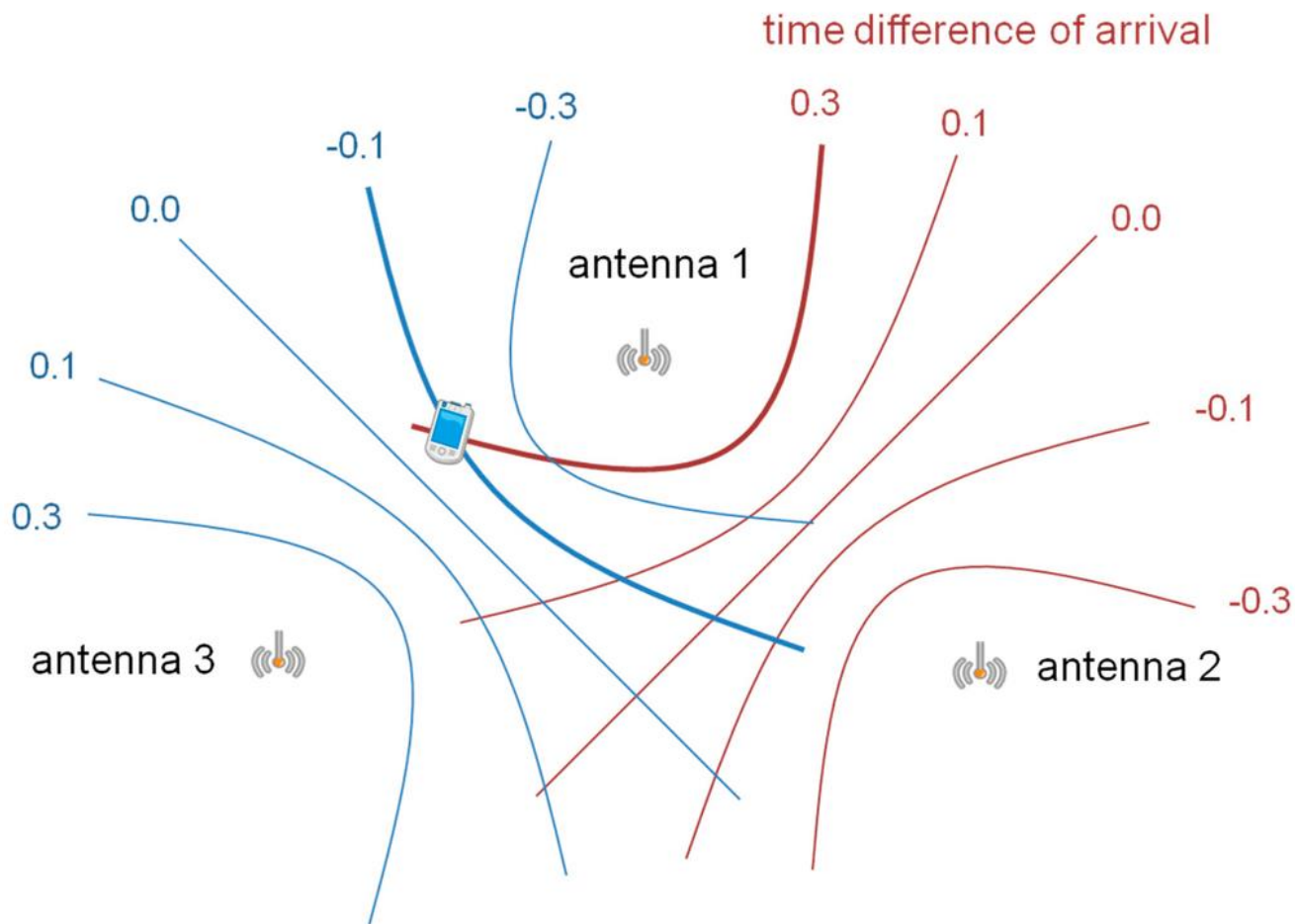
Source: The Wrongful Convictions Blog

Signal strength-based triangulation



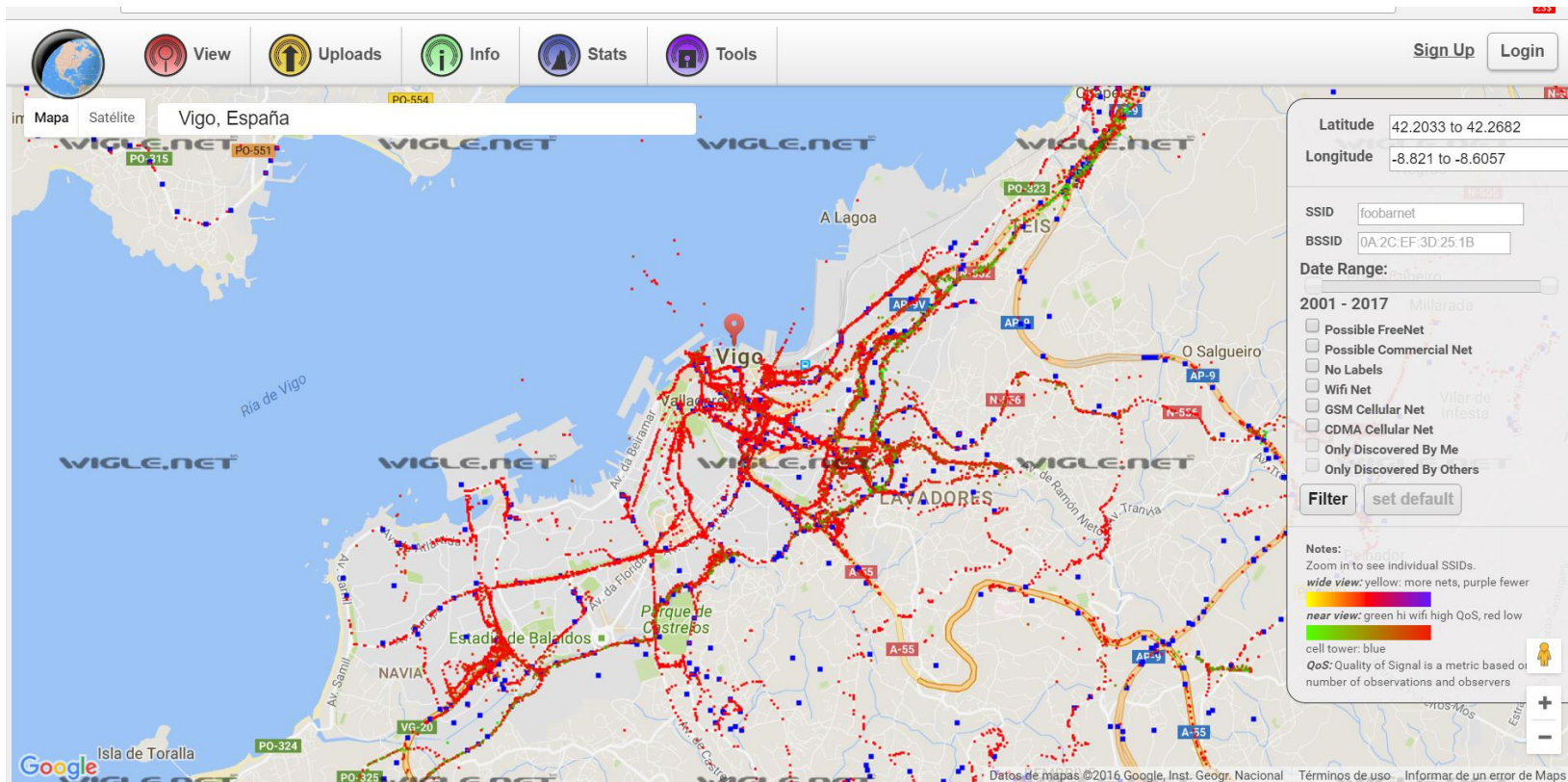
Source: The Wrongful Convictions Blog

Multilateration: Time Difference of Arrival (TDOA)



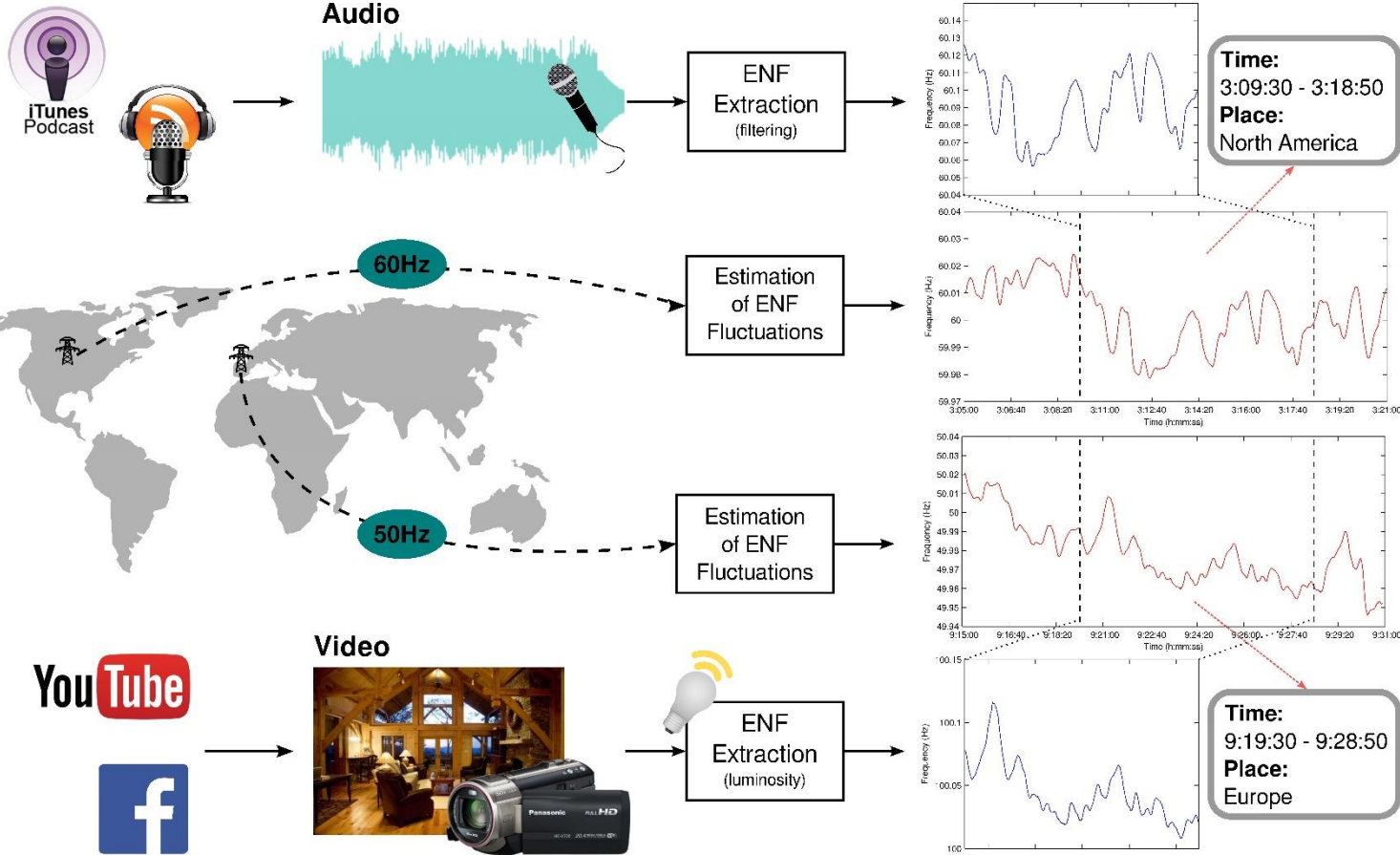
Source:[Fujii et al. 2015]

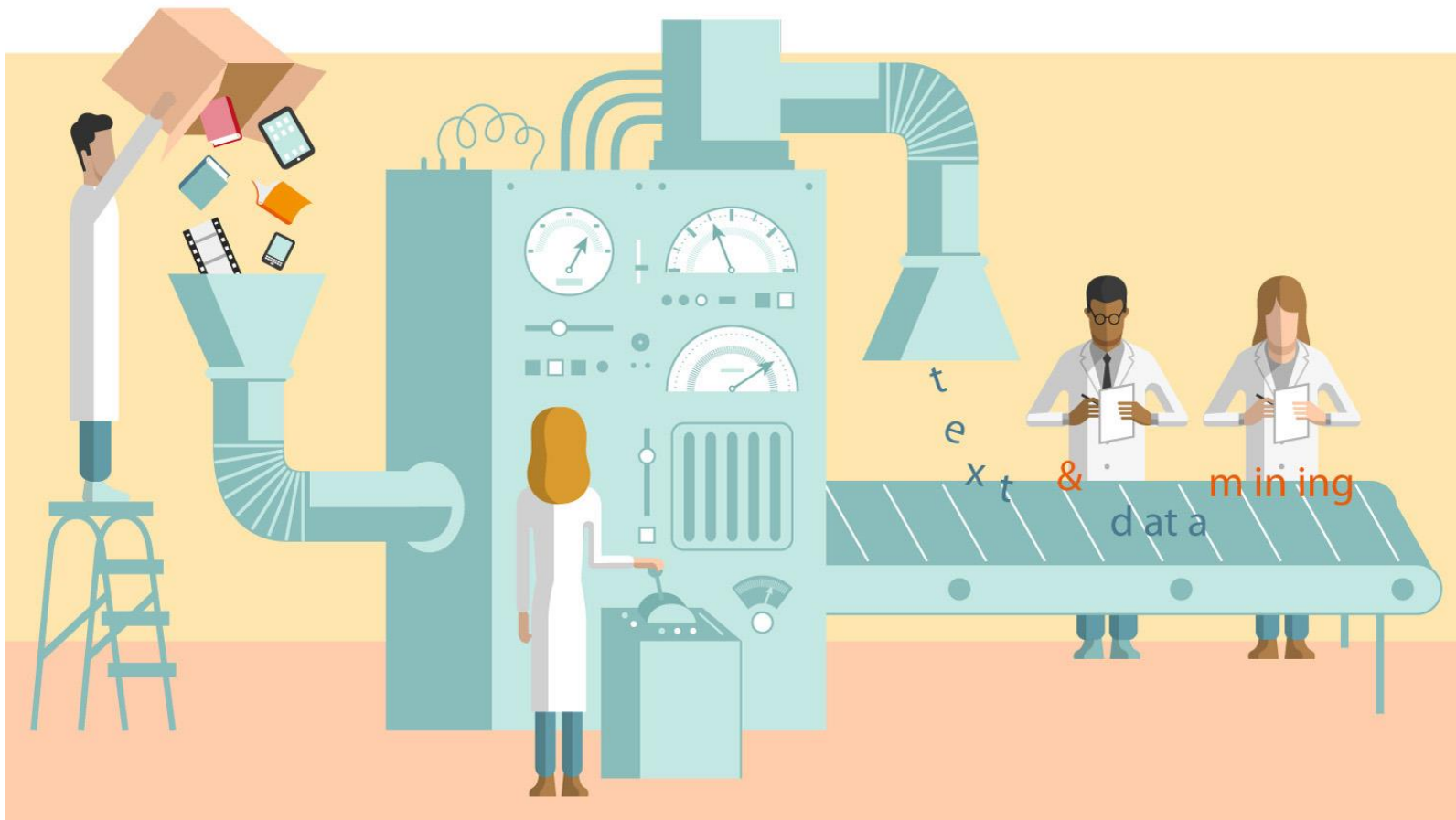
Wardriving geolocation (Wigle)

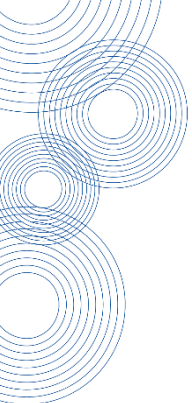


Source:Wigle.net

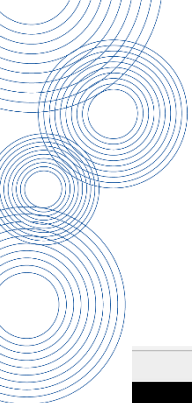
Electrical Network Frequency Geolocation







Why is it dangerous?



DOW JONES, A NEWS CORP COMPANY ▾

DJIA Futures ▲ 18116 0.10% Stoxx 600 ▼ 339.49 -0.38% U.S. 10 Yr ▲ 2/32 Yield 1.842% Crude Oil ▼ 48.49 -0.43% Euro ▼ 1.0956 -0.28%

THE WALL STREET JOURNAL.

Subscribe Now | Sign In

SPECIAL OFFER: JOIN NOW

Home World U.S. Politics Economy Business **Tech** Markets Opinion Arts Life Real Estate



Dyn Says Cyberattack Has Ended, Investigation Continues



Visa Taps Blockchain for Cross-Border Payment Plan



Airbnb Revises New York Rules Amid Possible Legislation



WHAT THEY KNOW

Websites Vary Prices, Deals Based on Users' Information

By JENNIFER VALENTINO-DEVRIES, JEREMY SINGER-VINE and ASHKAN SOLTANI
December 24, 2012

It was the same Swingline stapler, on the same [Staples.com](#) website. But for Kim Wamble, the price was \$15.79, while the price on Trude Frizzell's screen, just a few miles away, was \$14.29.

Staples seemed to think they were located

Most Popular Videos

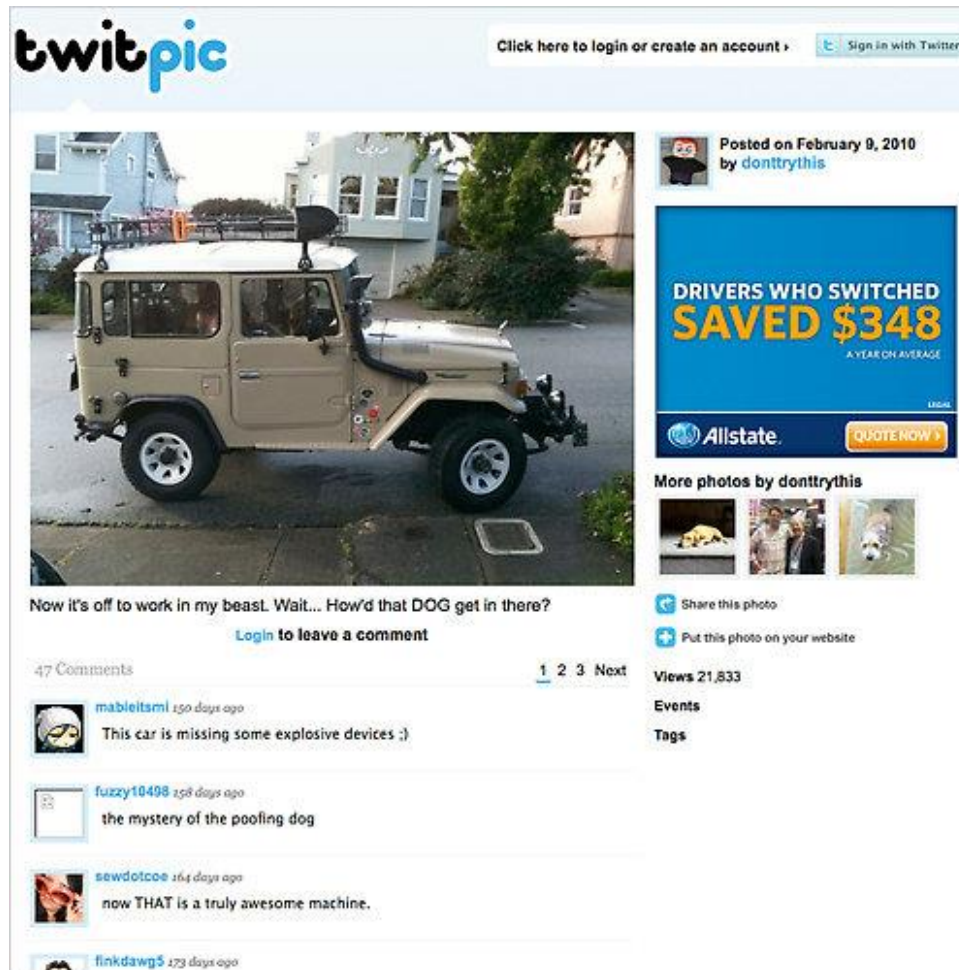
1. KLM to Make Final Dramatic Landing With 747



2. Bottle Flipping Hits a Wall



Buster busted!



The screenshot shows a twitpic post. At the top left is the 'twitpic' logo. To the right, there are links for 'Click here to login or create an account' and 'Sign in with Twitter'. The main image is a side view of a tan military-style jeep parked on a residential street. Below the image is the caption: 'Now it's off to work in my beast. Wait... How'd that DOG get in there?'. Underneath the caption is a 'Login to leave a comment' link. To the right of the image is a vertical advertisement for Allstate with the text 'DRIVERS WHO SWITCHED SAVED \$348 A YEAR ON AVERAGE' and a 'QUOTE NOW!' button. Below the ad are three small thumbnail images. Further down are social sharing options: 'Share this photo' and 'Put this photo on your website'. At the bottom right, it shows 'Views 21,833', 'Events', and 'Tags'. On the left side, below the caption, there are 47 comments. The first comment is from 'mabieitami' (150 days ago) saying 'This car is missing some explosive devices :)'. The second is from 'fuzzy10498' (158 days ago) saying 'the mystery of the pooling dog'. The third is from 'sewdotcoe' (164 days ago) saying 'now THAT is a truly awesome machine.'. The fourth is from 'finkdawg5' (173 days ago).



PLEASE ROB ME



Raising awareness about over-sharing

Check out our [guest blog post](#) on the CDT website.

Like Share 32K people like this. Be the first of your friends.

Check your own Twitter timeline for checkins

Are you curious if people can see your checkins?
Enter your Twitter username and find out.

Find!



More Info

[Home](#)

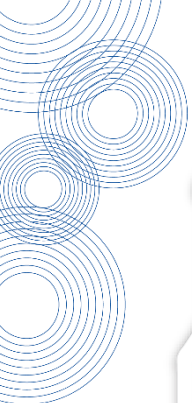
[Why](#)

Made Possible By

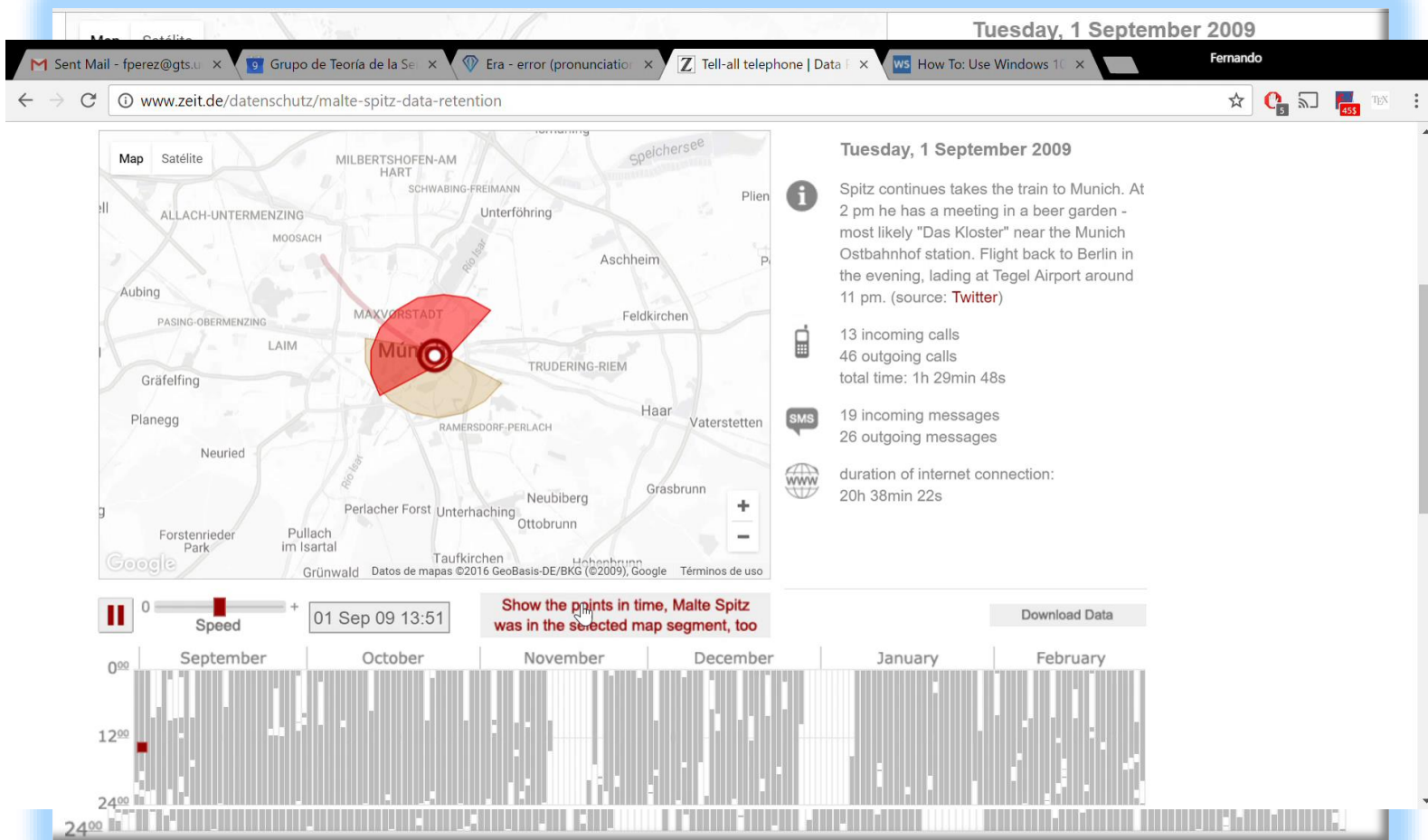
[Foursquare](#)

[Twitter](#)

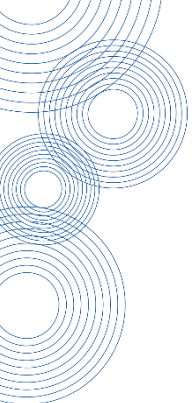
[@boyvanamstel](#)



6 months in the life of Malte Spitz (2009-2010)



Source:<http://www.zeit.de/datenschutz/malte-spitz-data-retention>

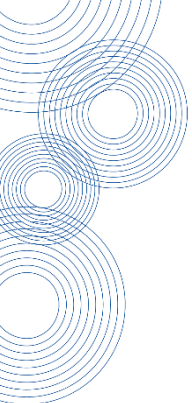


Are we concerned about it?

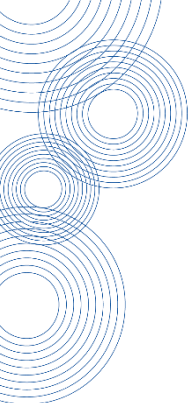
Are people really concerned about location privacy?

- Survey by Skyhook Wireless (July 2015) of 1,000 Smartphone app users.
- 40% hesitate or don't share location with apps.
- 20% turned off location for all their apps.
- Why people don't share location?
 - 50% privacy concerns.
 - 23% don't see value in location data.
 - 19% say it drains their battery.
- Why people turn off location?
 - 63% battery draining.
 - 45% privacy.
 - 20% avoid advertising.





How much is geolocation data worth?

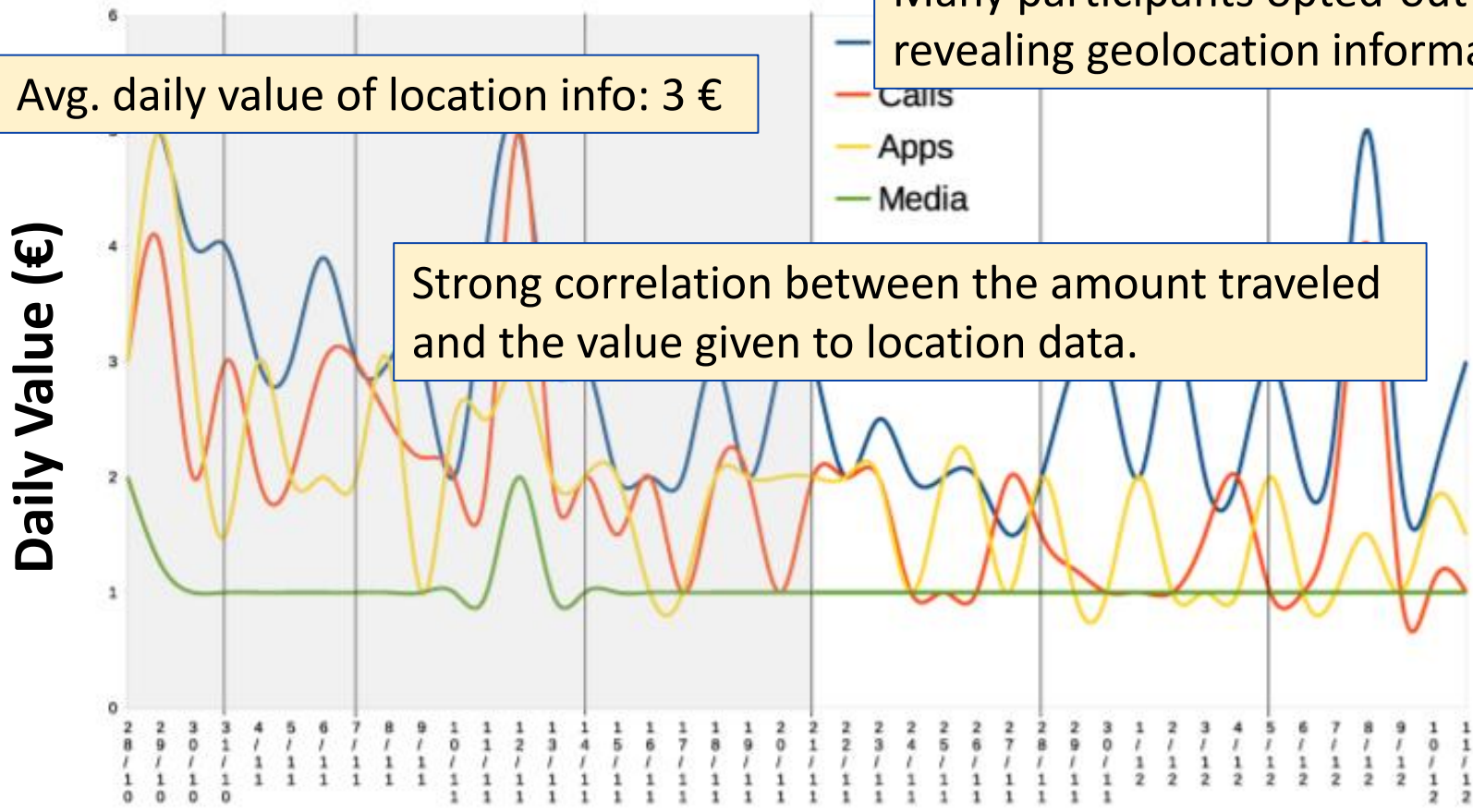


How much value do we give to location data? [Staiano et al. 2014]

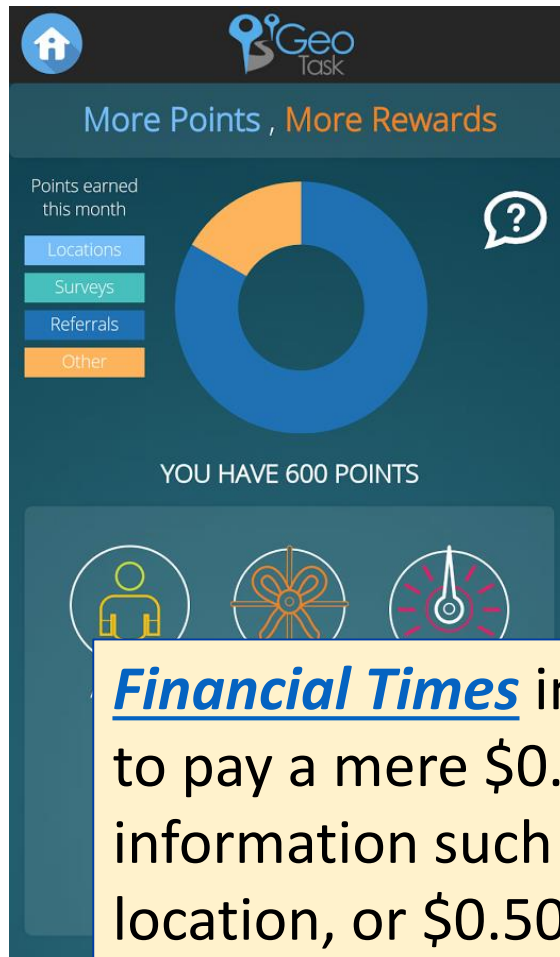
Avg. daily value of location info: 3 €

Many participants opted-out of revealing geolocation information.

Strong correlation between the amount traveled and the value given to location data.



Earn money as you share data



- GeoTask
- £1 PayPal cash voucher per 100 days of location data sharing (£0.01/day)

Financial Times in 2013: advertisers are willing to pay a mere \$0.0005 per person for general information such as their age, gender and location, or \$0.50 per 1,000 people.

Pay as you drive



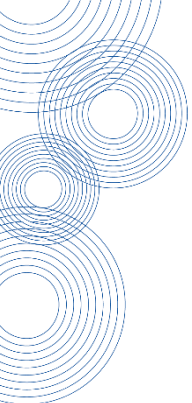
Home / Car Insurance / Pay As You Drive Insurance



Pay As You Drive Insurance

If you want the security of Comprehensive car insurance but you only drive a little, then

- Formula can be a function of the amount of miles driven, or the type of driving, age of the driver, type of roads used...
- Up to 40% reduction in the cost of insurance.



Home > Press Releases > Location-Targeted Mobile Advertising

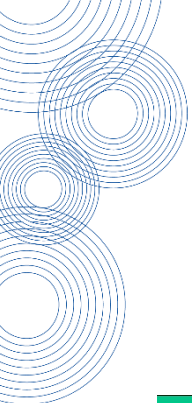


That's \$90 per person year!!!!

June 16, 2016 Press Releases

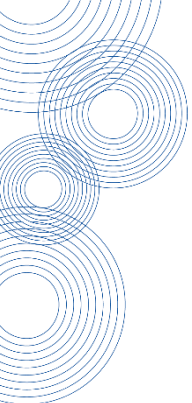
Native social advertising will represent more than one-quarter (28.1%) of U.S. local-targeted ad spend by 2020, pulling market share from search and display.

BIA/Kelsey projects U.S. location-targeted mobile ad spending to grow from \$9.8 billion in 2015 to \$29.5 billion in 2020.



SAP, Germany, estimates wireless carrier revenue from selling mobile-user behavior data in \$5.5 billion in 2015 and predicts \$9.6 billion for 2016.

Symbol	Price	Change
DOW	142.45	-1.32
Verizon	146.96	-1.21
Disney	88.24	-1.43



How about anonymization/pseudonymization?

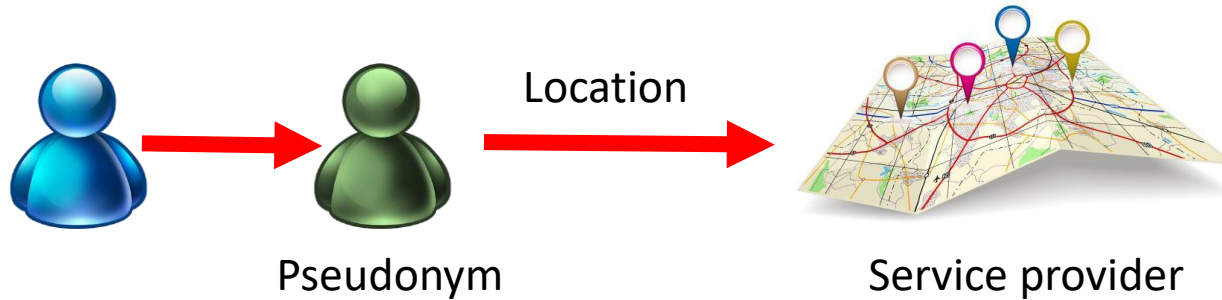
Anonymity



Problems:

- Difficult authentication and personalization.
- Operating system or apps may access location before anonymization.

Pseudonymity

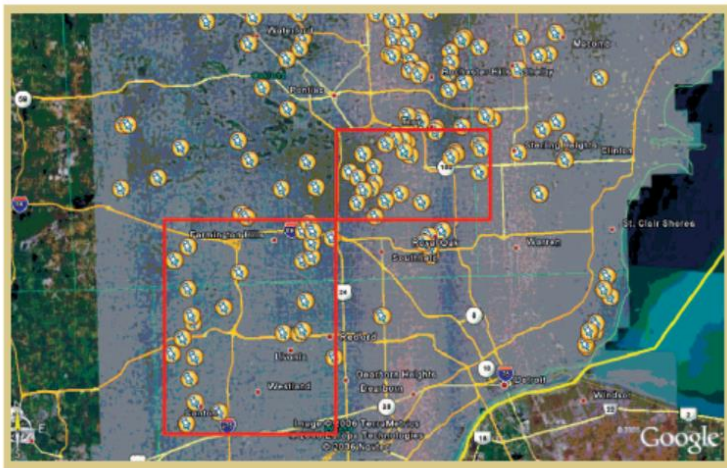


Problems:

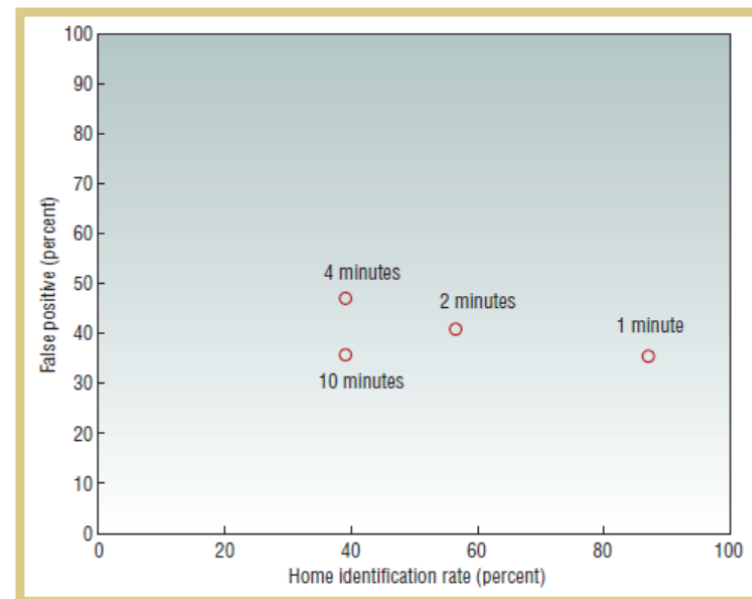
- Operating system or apps may access location data before pseudonymization.
- Deanonimization.

Deanonymization based on home location [Hoh, Gruteser 2006]

- Data from GPS traces of larger Detroit area (1 min resolution).
- No data when vehicle parked.
- K-means algorithm for clustering locations + 2 heuristics:
 - Eliminate centroids that don't have evening visits.
 - Eliminate centroids outside residential areas (manually).



Source: [Hoh, Gruteser 2006]

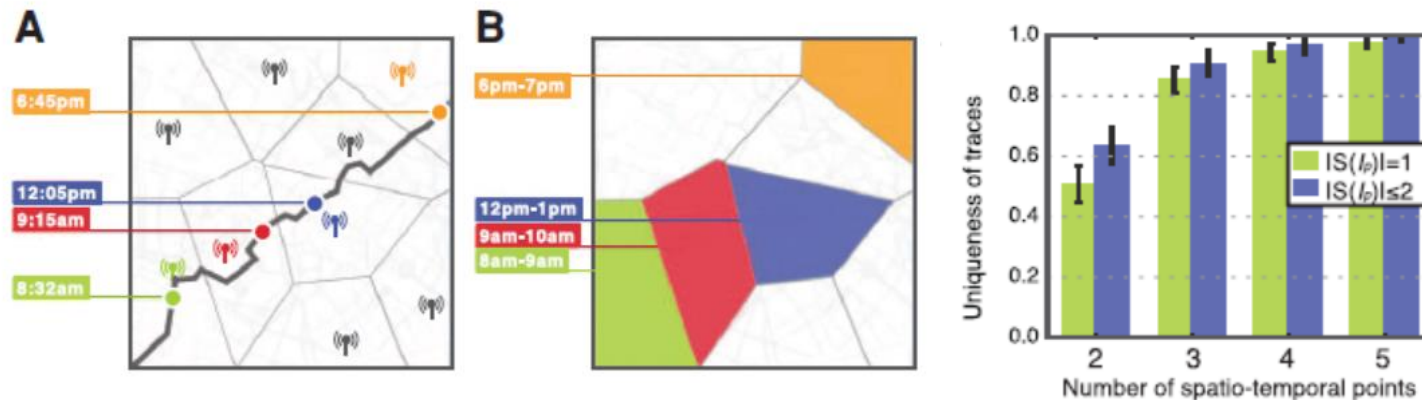


Deanonymization based on home location [Krummer 2007]

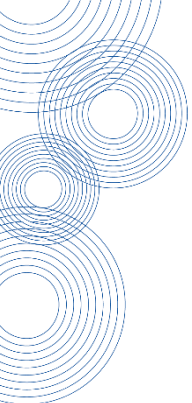
- 2- week GPS data from 172 subjects (avg. 6 sec resolution).
- Use heuristic to single out trips by car.
- Then use several heuristics: destination closest to 3 a.m. is home; place where individual spends most time is home; center of cluster with most points is home.
- Use reverse geocoding and white pages to deanonymize. Success measured by finding out name of individual.
- Positive identification rates around 5%.
- Even noise addition with std=500 m gives around 5% success when measured by finding out correct address.

Mobile trace uniqueness [de Montjoye et al 2013]

- Study on 15 months of mobility data; 0.5M individuals.
- Dataset with hourly updates and resolution given by cell carrier antennas, only 4 points suffice to identify 95% of individuals.
- Uniqueness of mobility traces decays as 1/10th power of their resolution.



Source: [de Montjoye et al. 2013]



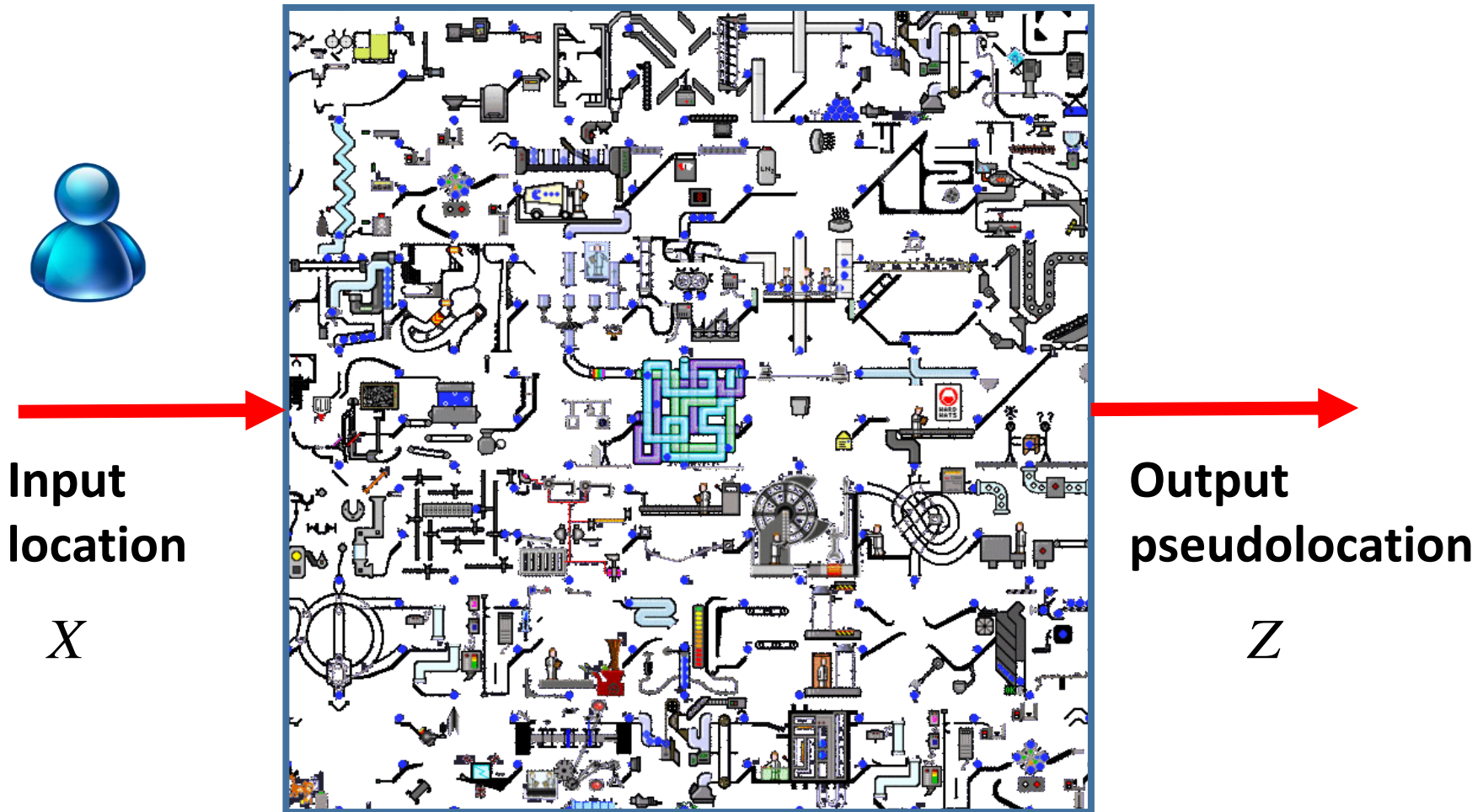
Location privacy protection mechanisms

Location white lies



Source: Caro Spark (CC BY-NC-ND)

Location based privacy mechanisms



Source: Motherboards.org

Location privacy protection mechanisms (LPPMs)

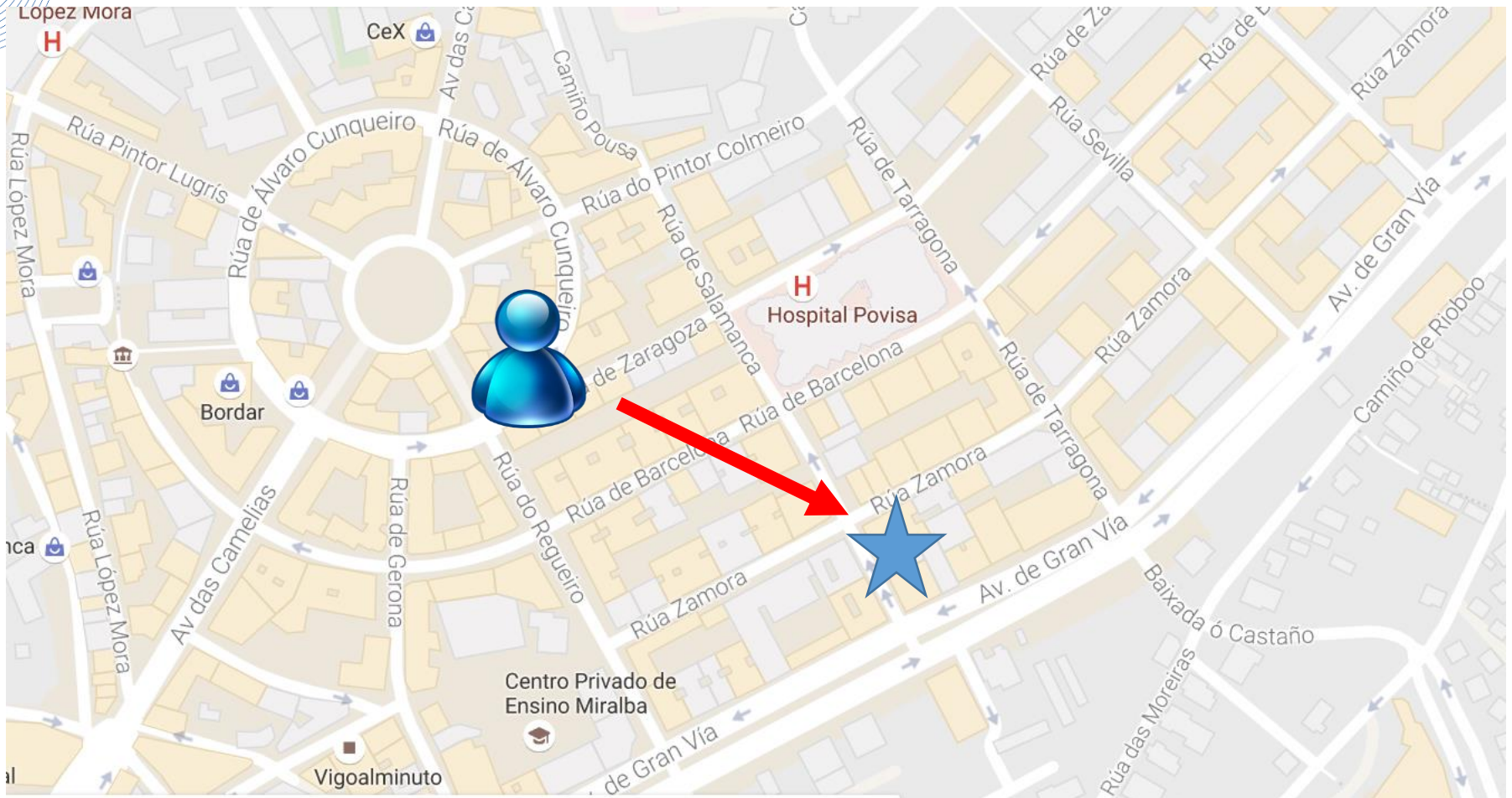
- $Z = \varphi(X)$
- The **mechanism** may be deterministic (e.g., quantization) or stochastic (e.g., noise addition).
- Function $\varphi(\cdot)$ may depend on other contextual (e.g., time) or user-tunable (e.g., privacy level) parameters.
- When the **mechanism is stochastic**, there is an underlying probability density function, i.e.,

$$f(Z | X)$$

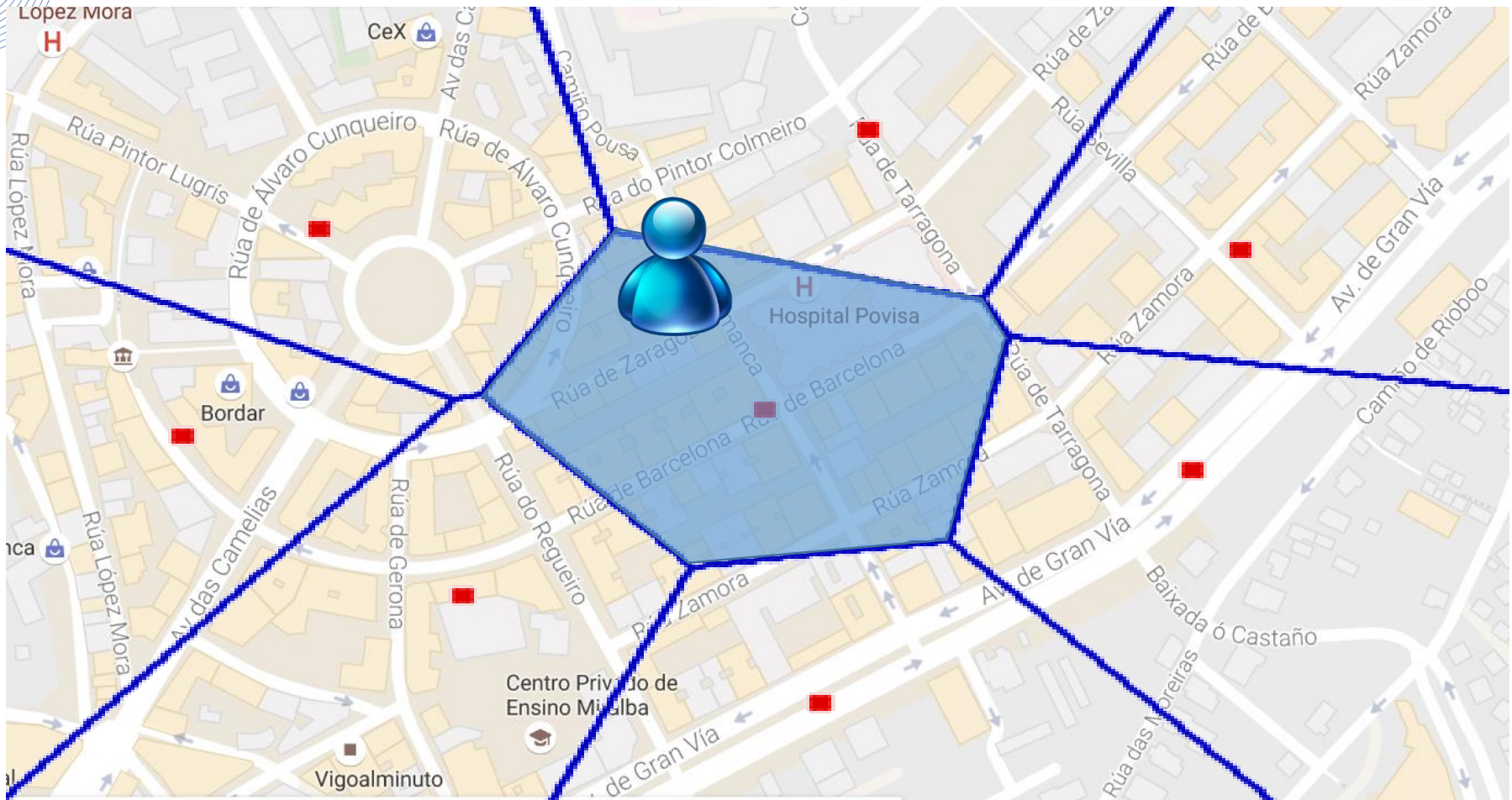
Hiding



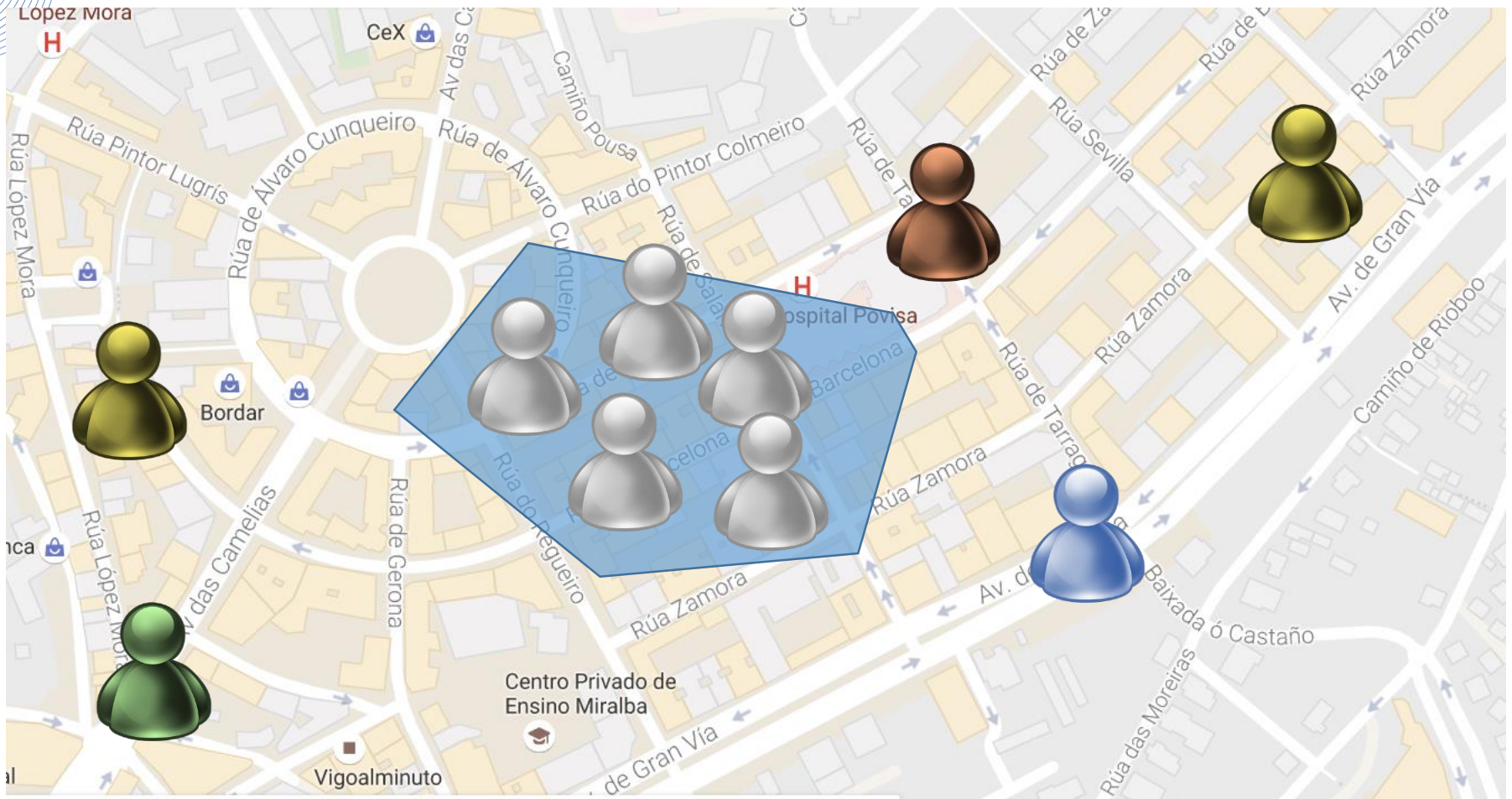
Perturbation: (independed) noise addition



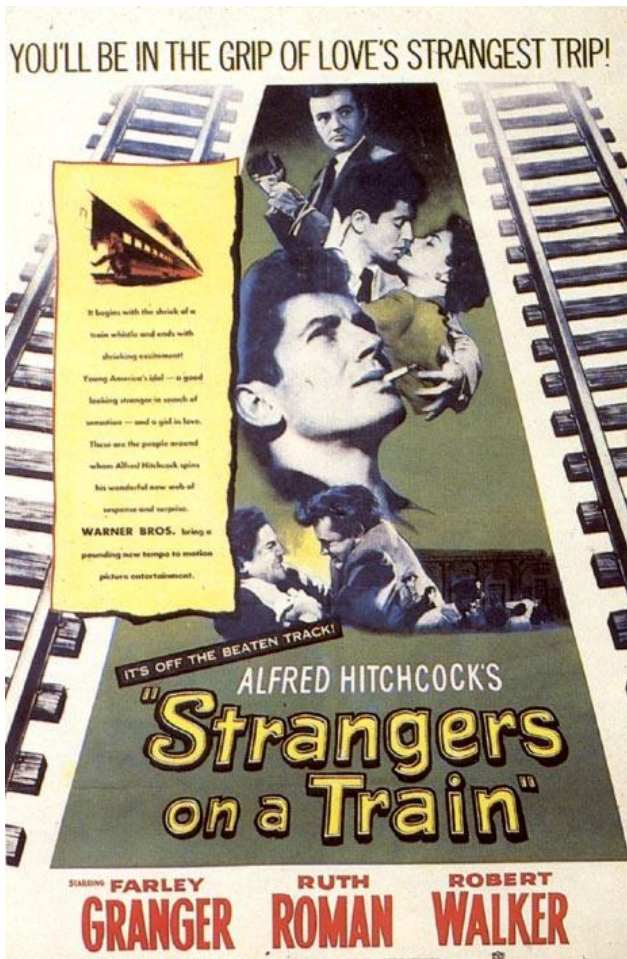
Obfuscation



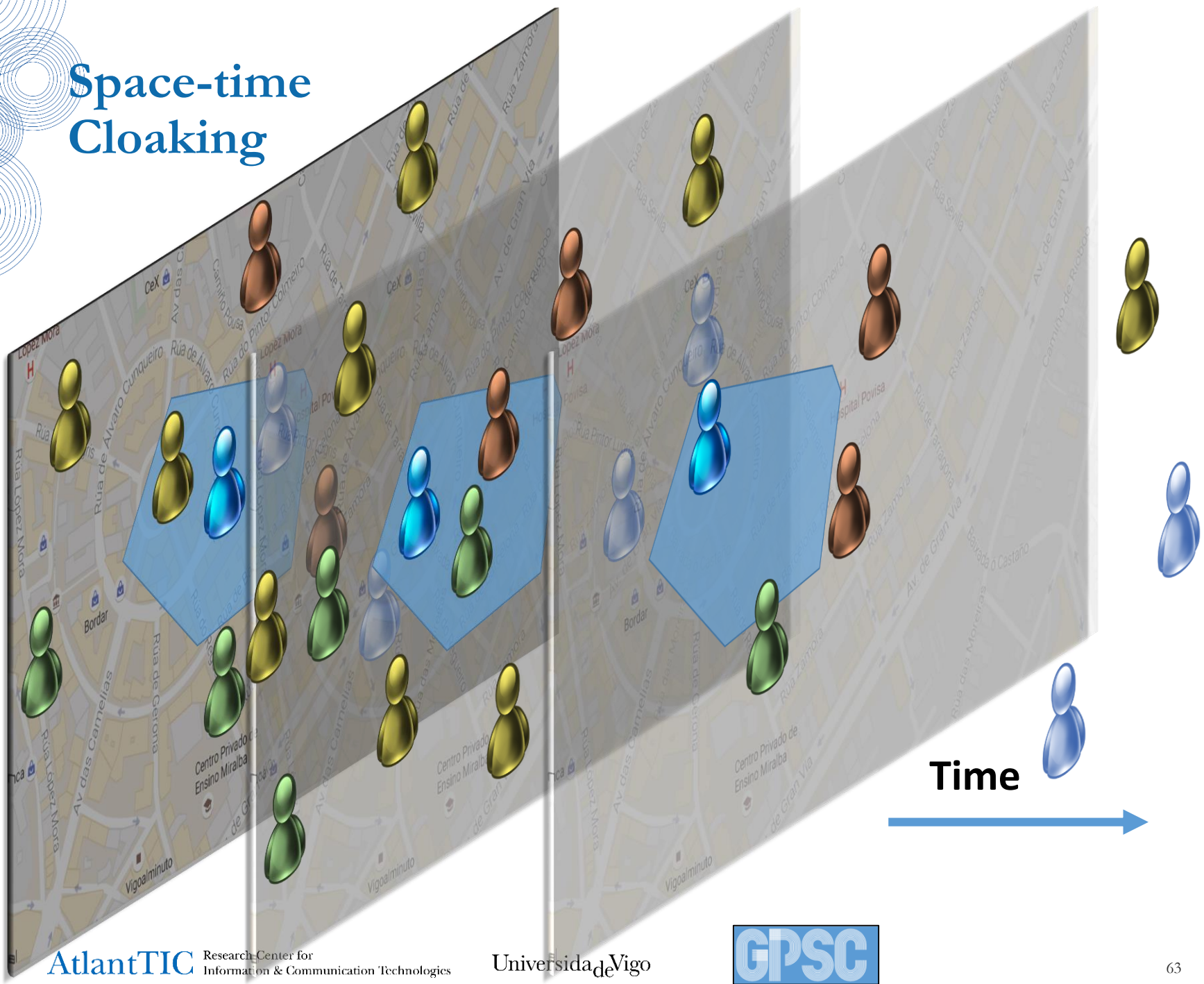
Spatial Cloaking



How to commit the perfect murder



Space-time Cloaking



Dummies

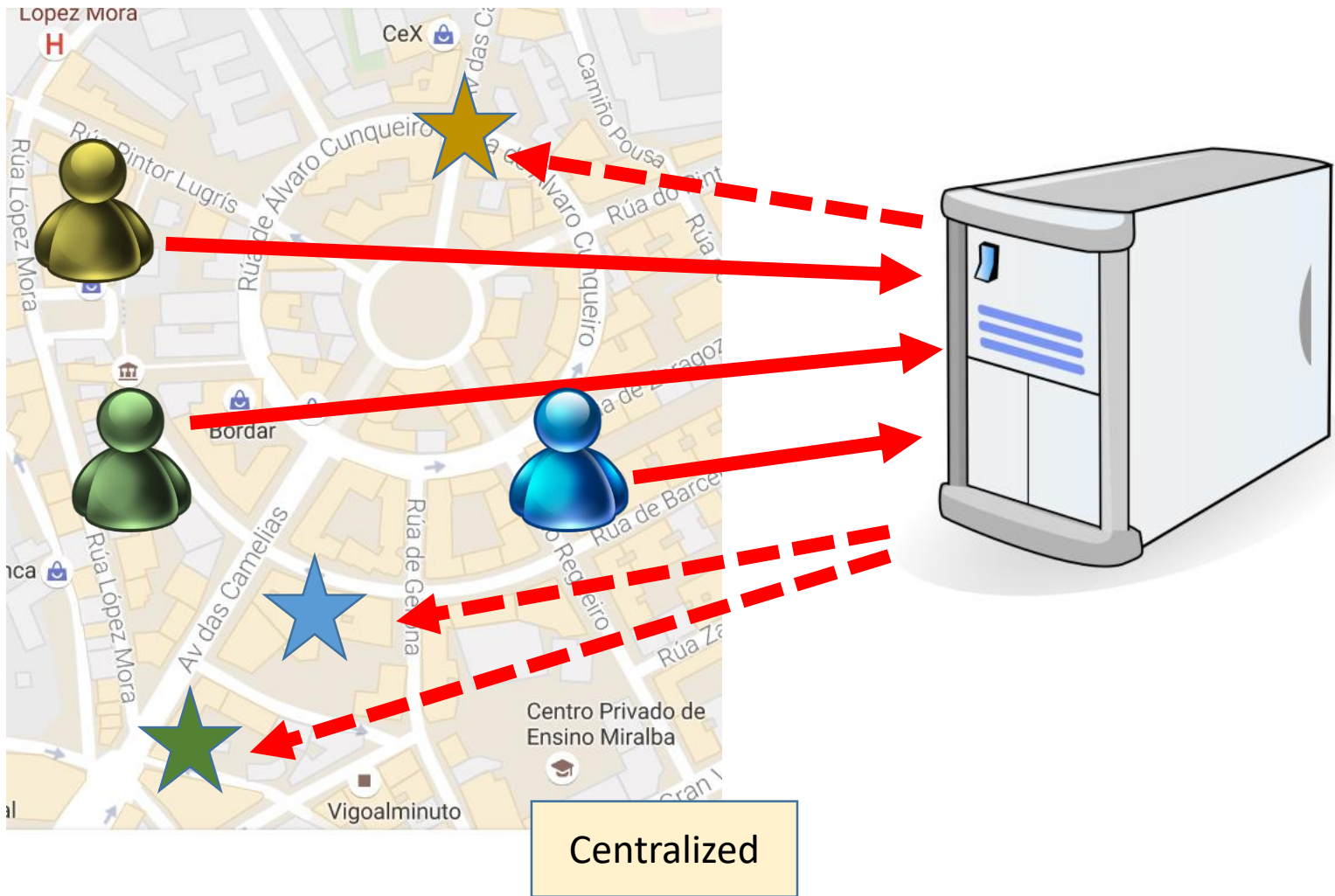


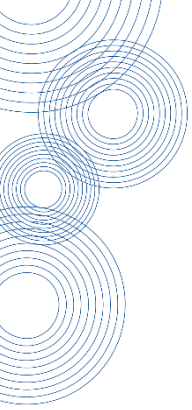
User-centric vs. Centralized LPPM



User-centric

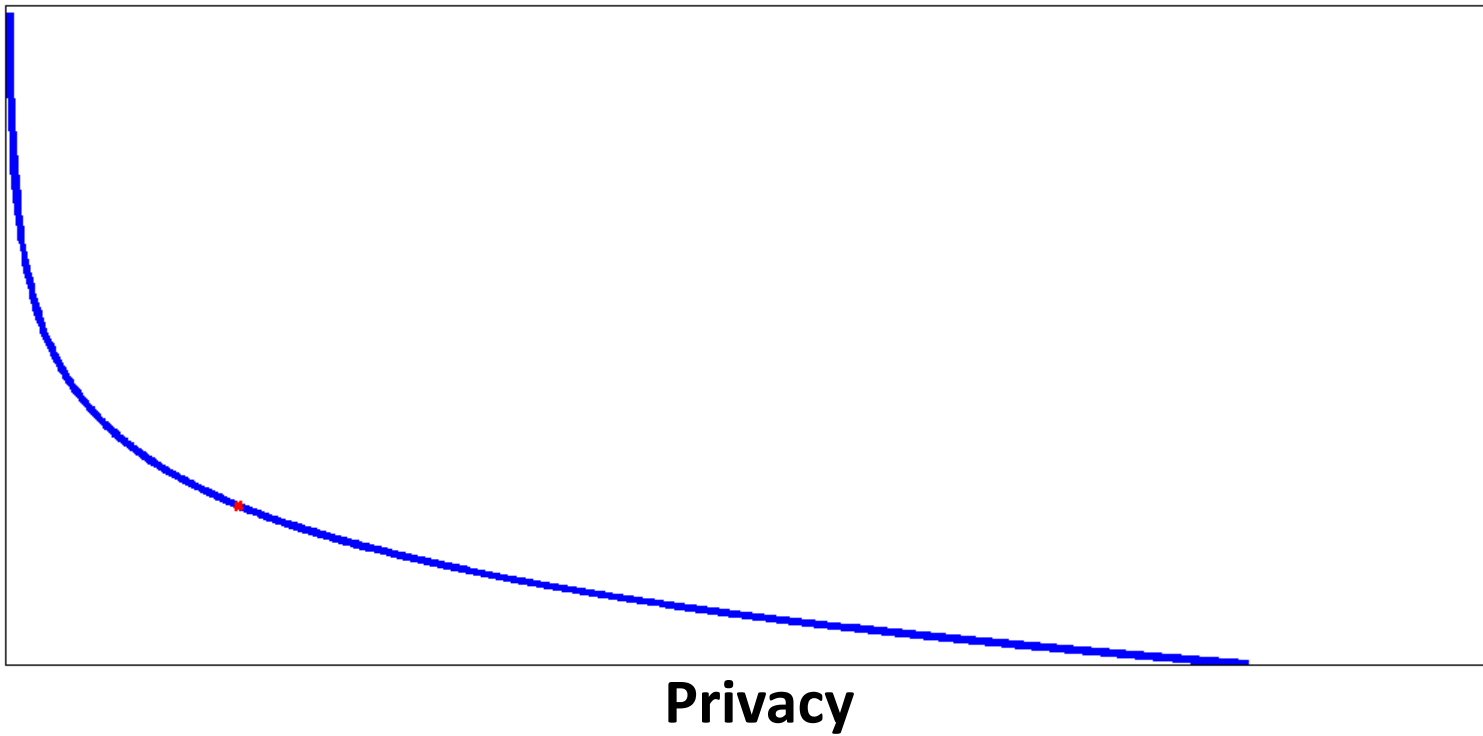
User-centric vs. Centralized LPPM





Utility vs. Privacy

- In broad terms:

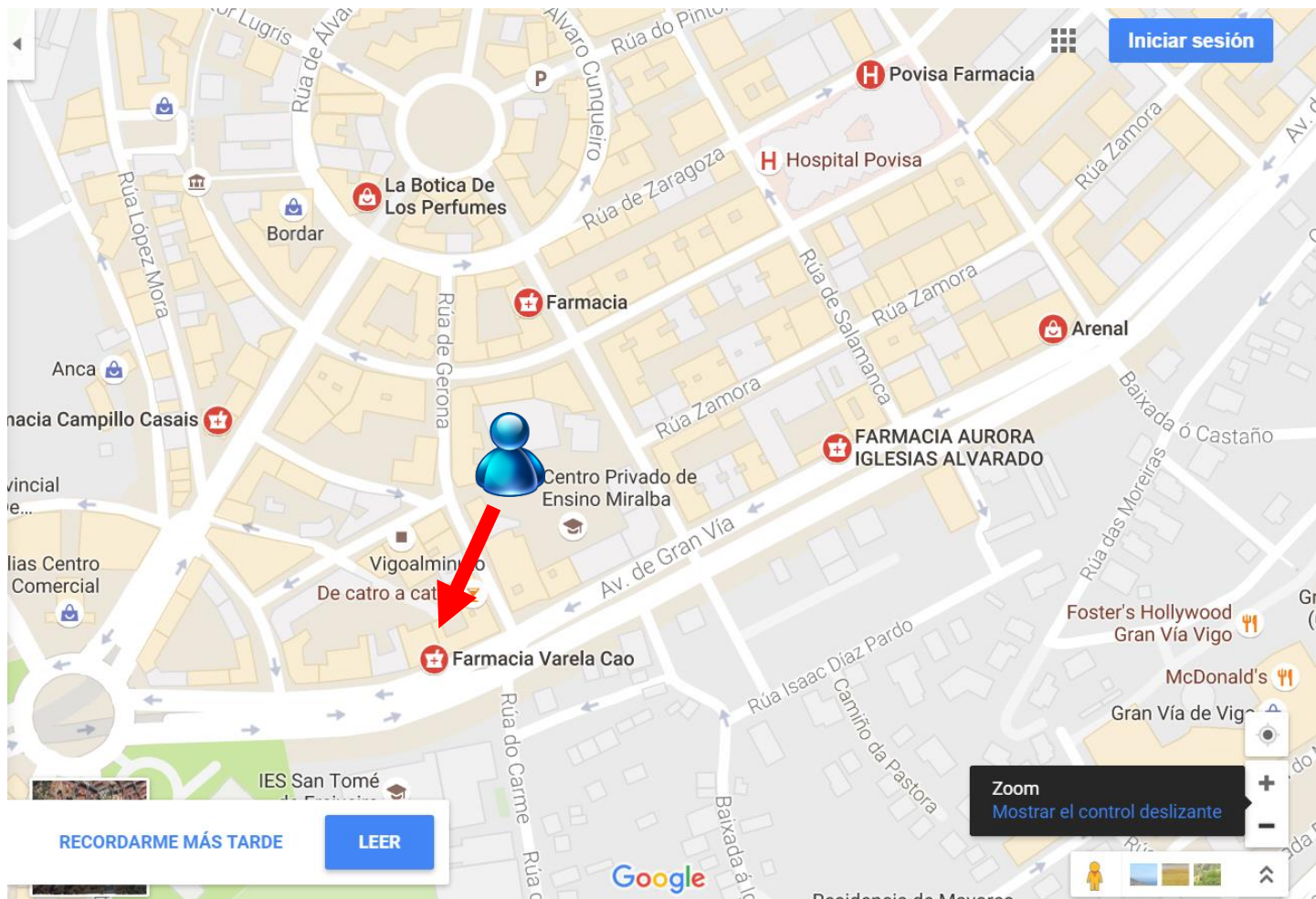


Very nice, but...

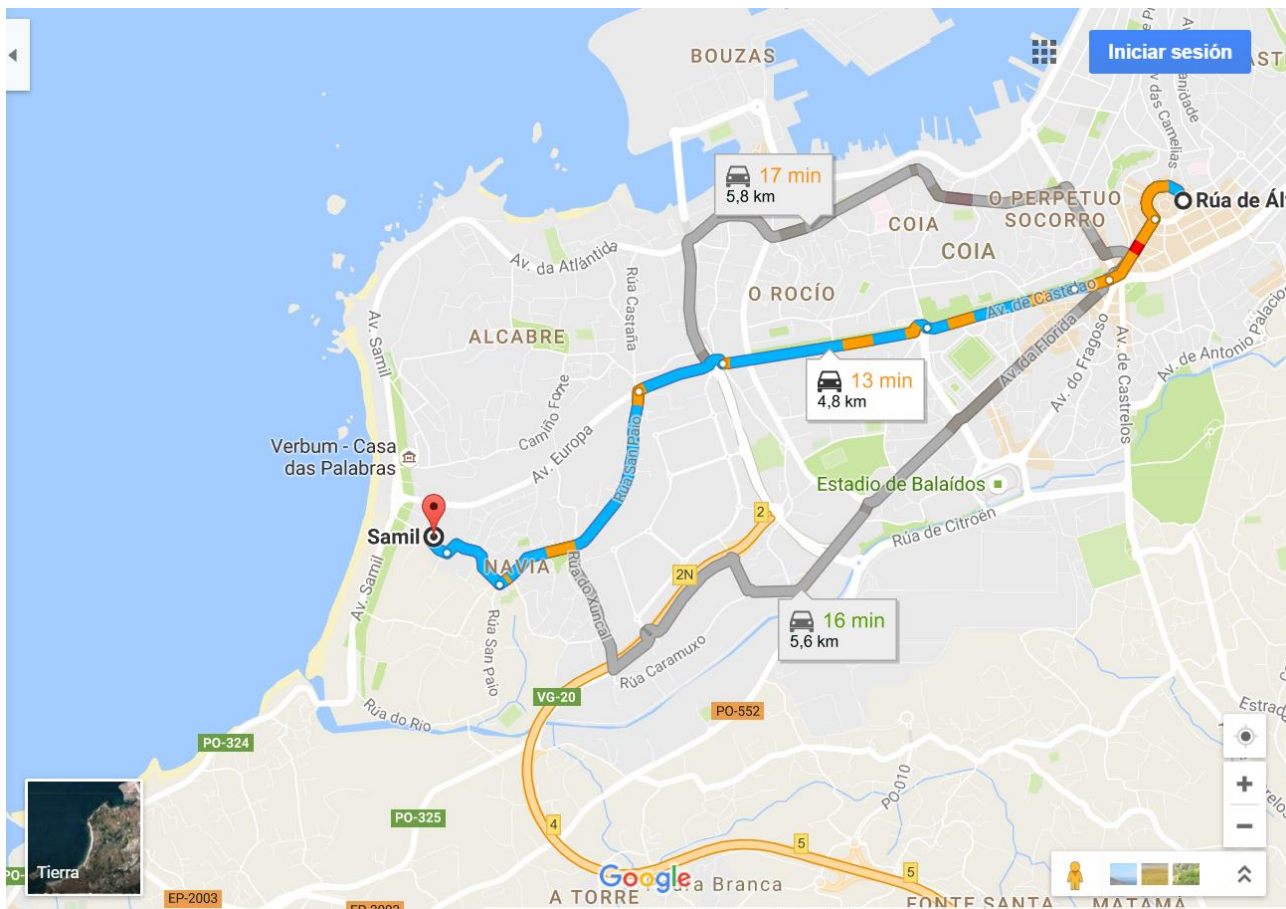
- There are two main problems:
How do we measure utility?
How do we measure privacy?



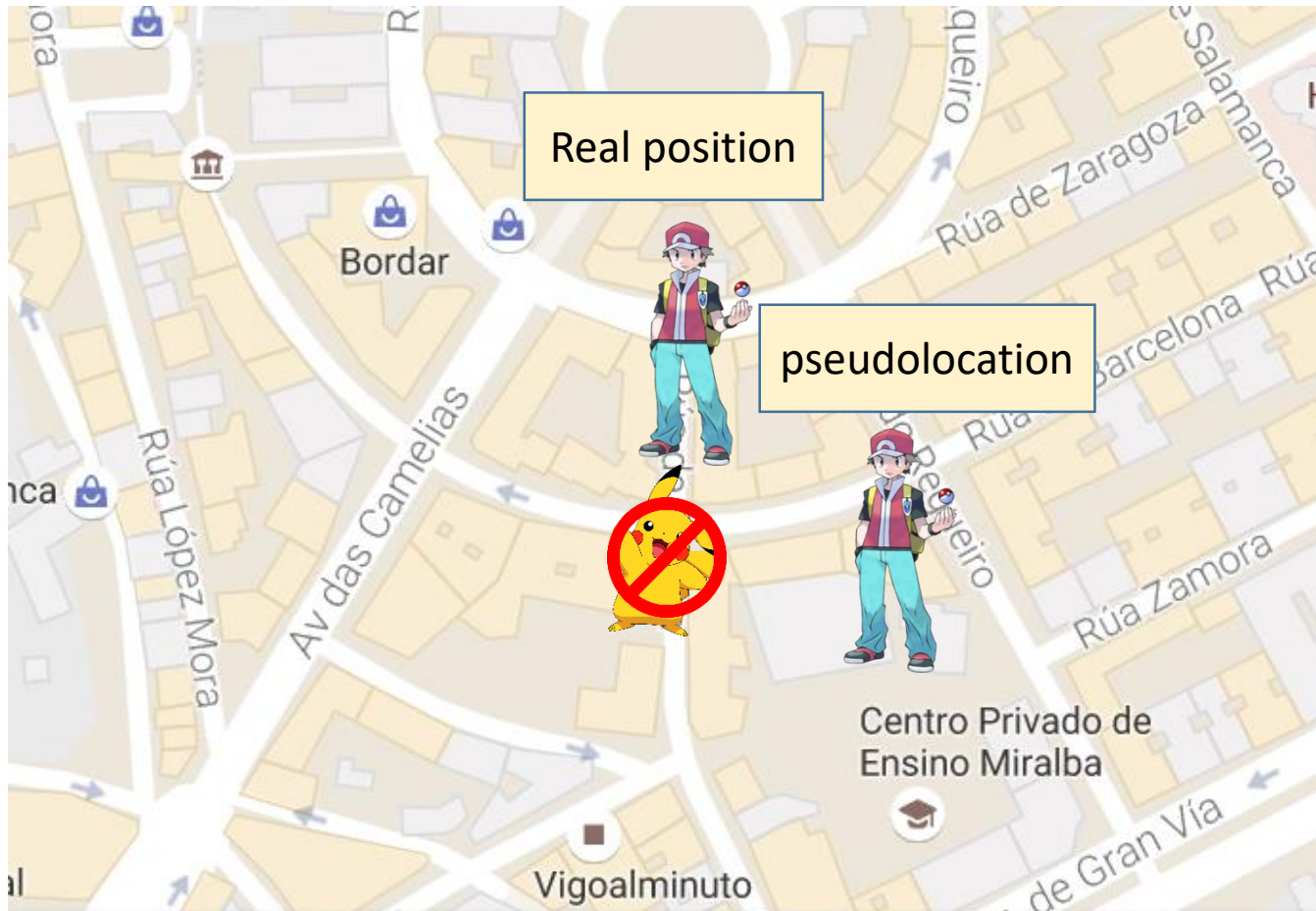
How to measure utility?



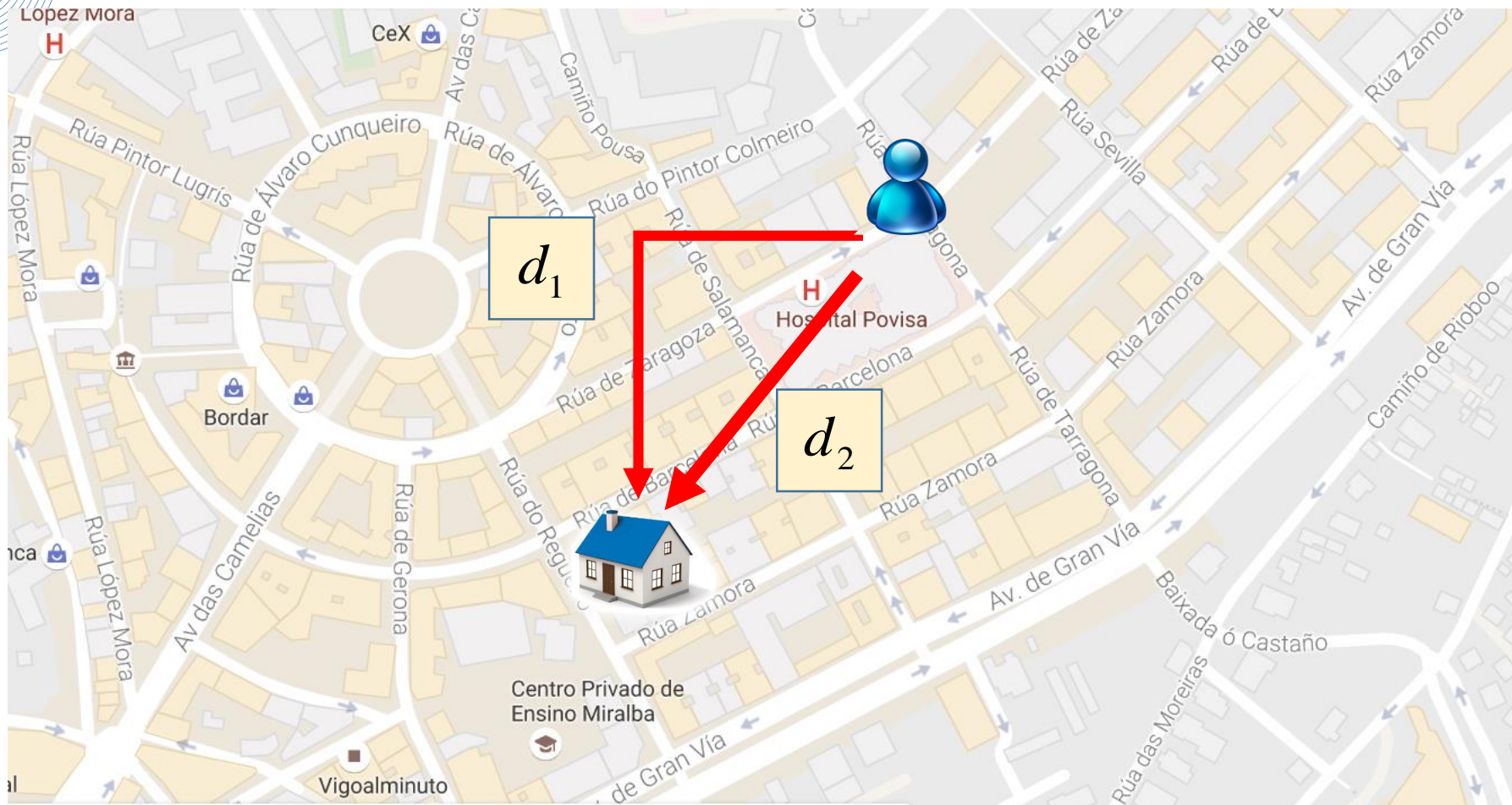
How to measure utility?



How to measure utility?



A note about distances



Adversarial definition of privacy [Shokri et al 2011-]

- Assume stochastic mechanism for the user $f(Z | X)$.
- Adversary constructs a (possibly stochastic) estimation remapping $r(\hat{X} | Z)$.
- Prior $\pi(X)$ assumed available to the adversary.
- $d_p(x, \hat{x})$: Distance between \hat{x} and x .
- $d_q(x, z)$: Distance between x and z .



Adversarial definition of privacy [Shokri et al 2011-]

- Establish a cap on average utility loss: $E\{d_q(X, Z)\} \leq QL$
- This is a Stackelberg game in which the user chooses first and the adversary plays second.
- Find optimal adversarial 'remapping':

$$r^*(\hat{X} | Z) = \arg \min E\{d_p(\hat{X}, X) | Z\}$$

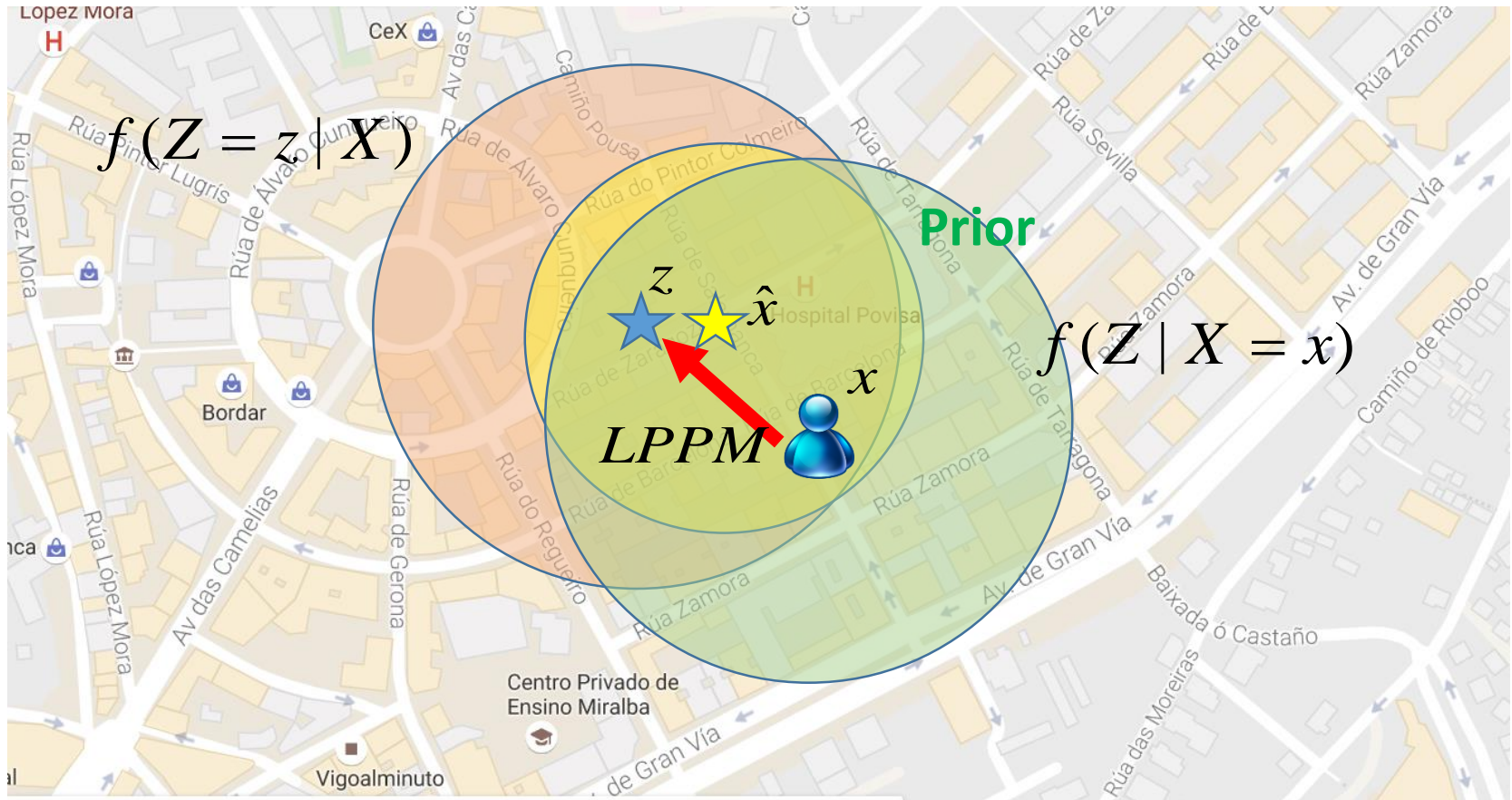
- Optimal remapping depends on $f(Z | X)$ and $\pi(X)$.

$$E\{d_p(\hat{X}, X) | Z\} = \sum_{X, \hat{X}} r(\hat{X} | Z) f(X | Z) d_p(\hat{X}, X)$$

where

$$f(X | Z) = \frac{f(Z | X) \cdot \pi(X)}{f(Z)}$$

Example: uniform noise addition



Adversarial definition of privacy [Shokri et al 2011-]

- When for a given Z there are several minimizers \hat{X} the function $r^*(\hat{X} | Z)$ becomes stochastic.
- The user now must maximize privacy:

$$\max E\{d_p(\hat{X}, X)\} = \max \sum_{Z, X, \hat{X}} r^*(\hat{X} | Z) f(Z | X) \pi(X) d_p(X, \hat{X})$$

- Which is achieved for some mechanism $f^*(Z | X)$
- Privacy is defined as $E\{d_p(\hat{X}, X)\}$ after solving this maxmin problem.

An interesting result

- When $d_p = d_q$:

$$f^*(Z = z | X) = \arg \min E\{d_p(z, X)\}$$

$$r^*(\hat{X} | Z = z) = \delta(\hat{X} - z)$$

i.e. do nothing!

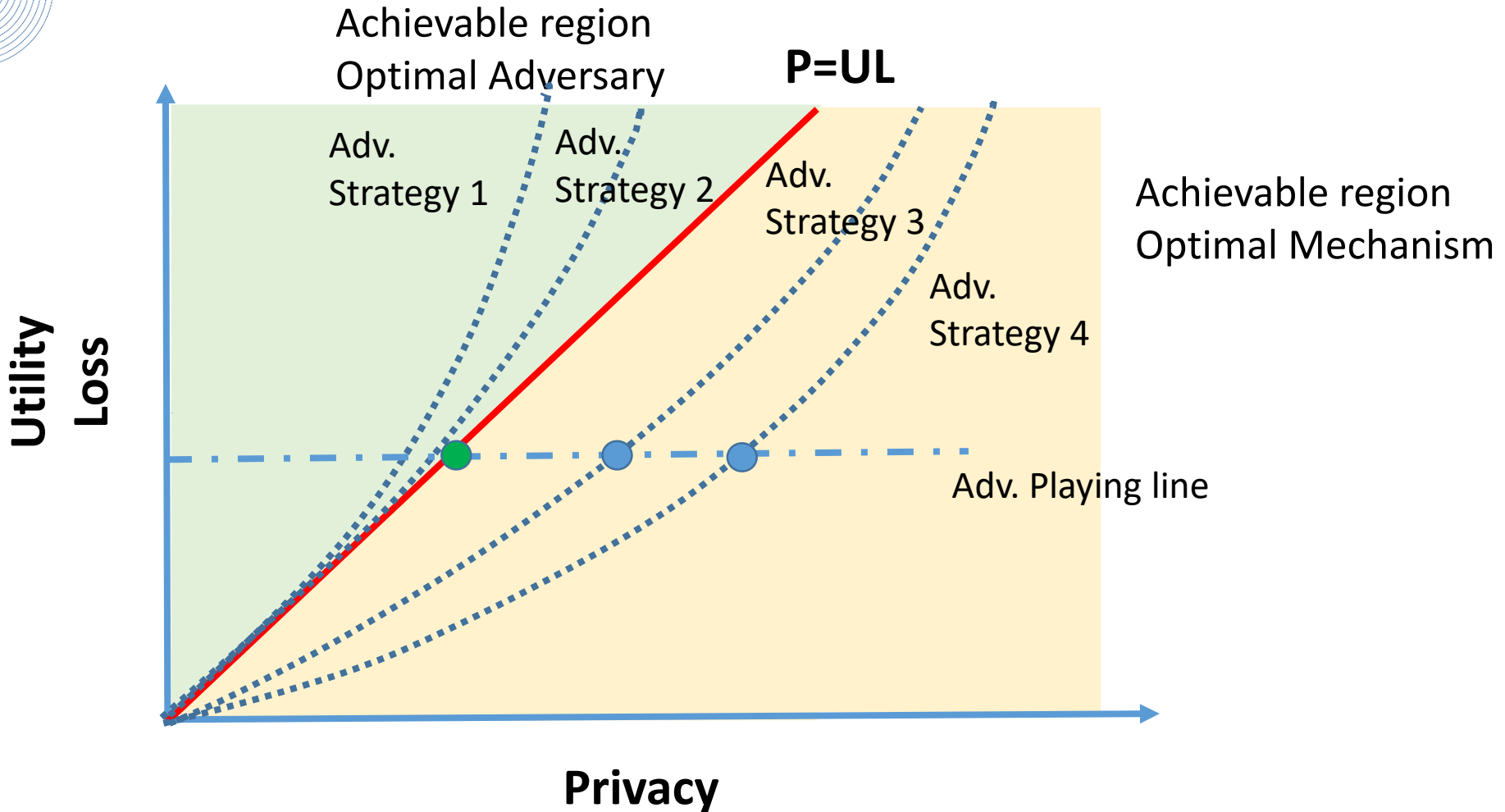
- When $d_p = d_q = d_2$ the following identity must hold

$$z = E\{X | Z = z\}$$

- When both user and adversary play optimally:

Privacy=Utility Loss

The Utility Loss-Privacy plane



What's wrong with priors?

- Is it realistic to assume that the adversary knows the prior?
- Adversary no longer plays optimally with the 'wrong' prior.
- Shokri's privacy definition is prior-dependent.
- Definition of differential privacy is prior-independent:

$$\log(\Pr\{A(D_1) \in S\}) \leq \varepsilon + \log(\Pr\{A(D_2) \in S\})$$

- Two databases D_1, D_2 differing in a single element.
- A : randomized algorithm.
- S : set of possible subsets of $im(A)$.

Geoindistinguishability [Chatzikokolakis et al 2013-]

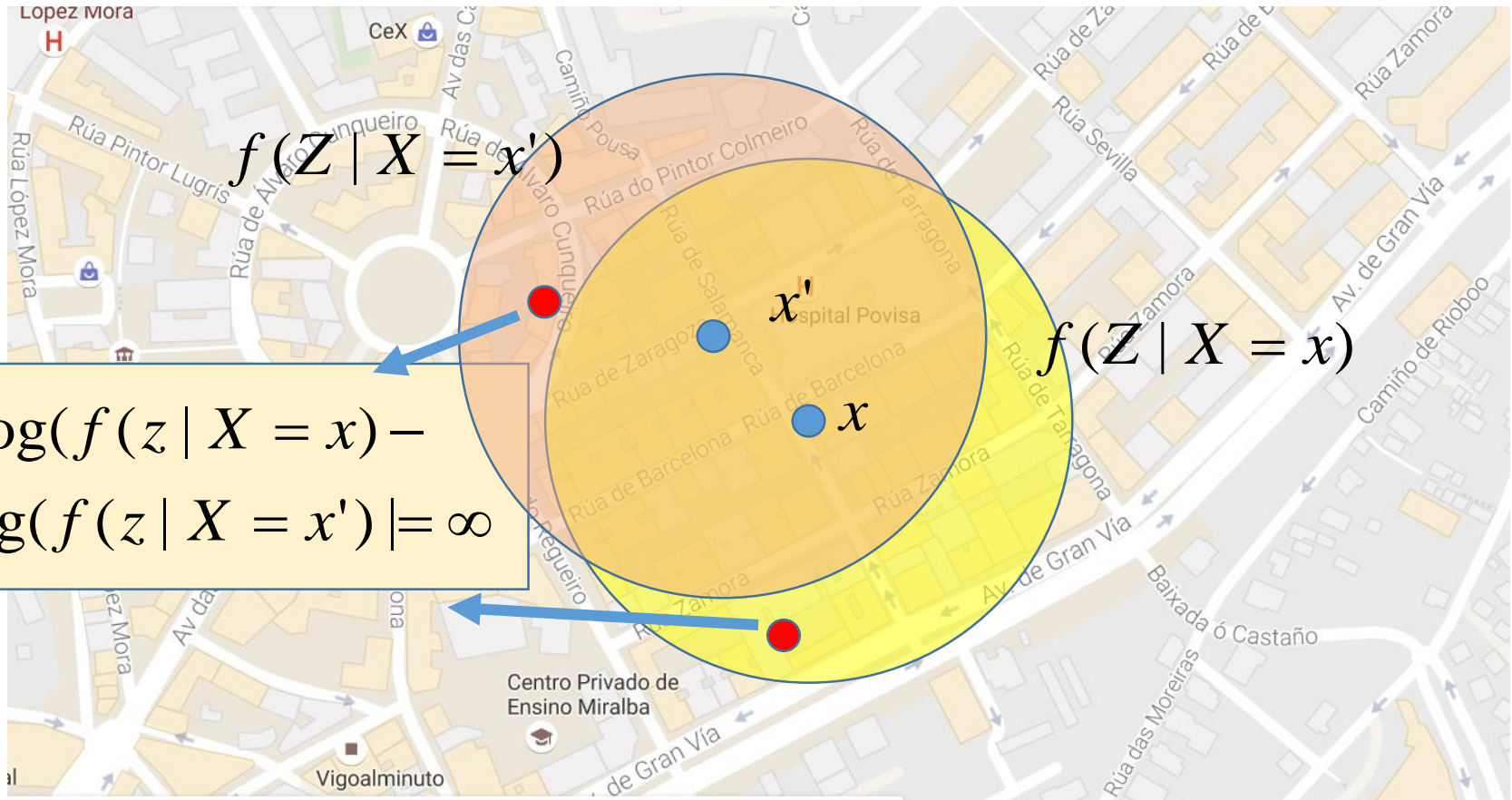
- A mechanism is ε -geo-indistinguishable iff:

$$|\log(f(z | X = x)) - \log(f(z | X = x'))| \leq \varepsilon \cdot d_p(x, x')$$

for all x, x', z .

- Differential privacy corresponds to $d_p =$ Hamming distance.
- Definition is **prior-independent**.
- Guarantees a small leakage of information BUT is no defense against EVERY adversary: with proper side information, adversary can learn a lot!

Uniform mechanisms do not provide geo-ind



Laplacian mechanism

- Laplacian distribution in polar coordinates:

$$f(z | X = x) = \frac{\varepsilon}{2\pi} e^{-\varepsilon \cdot d_2(x,z)}$$

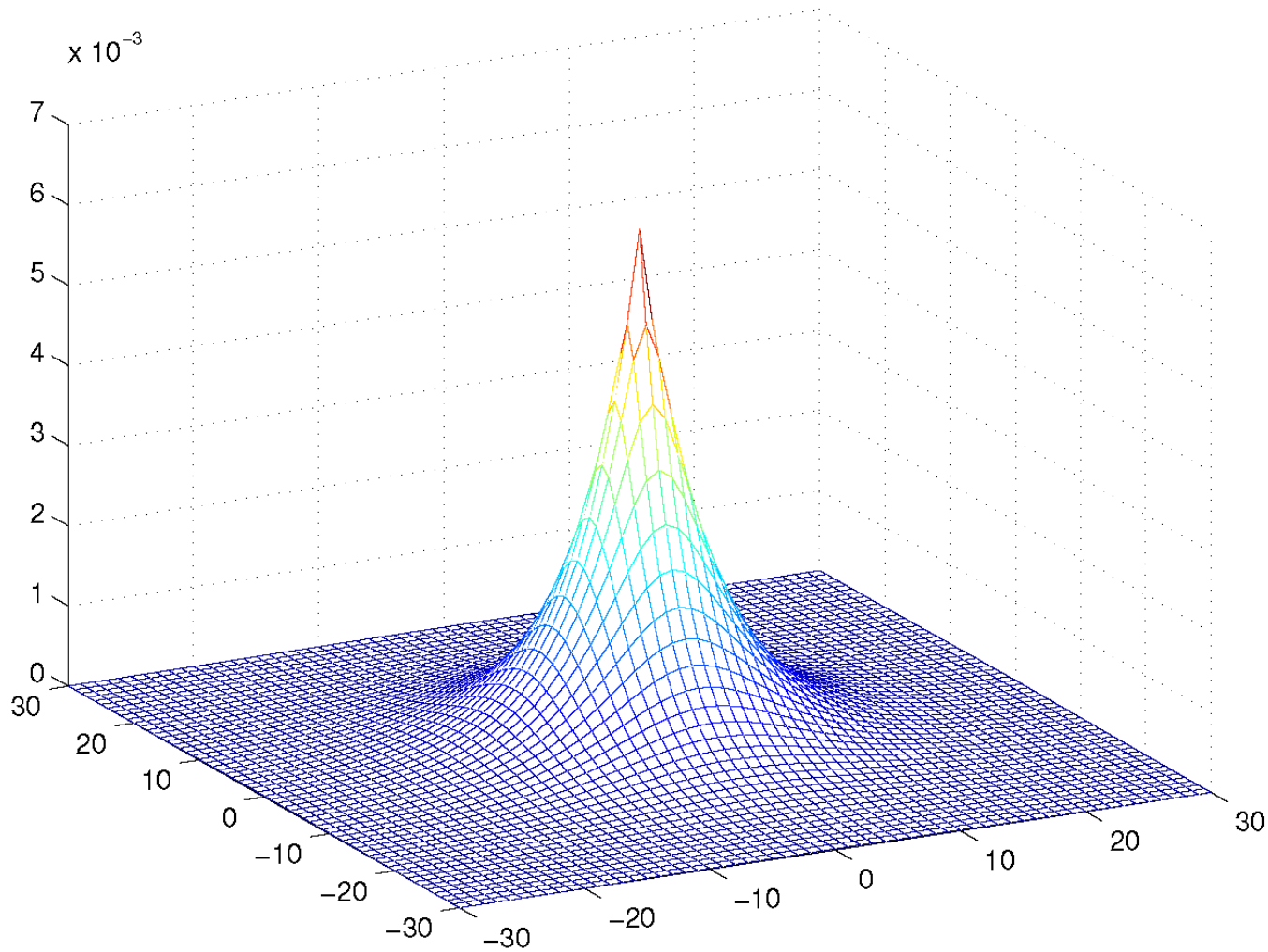
- Then,

$$|\log f(z | X = x) - \log f(z | X = x')| = |\varepsilon \cdot d_2(z, x') - \varepsilon \cdot d_2(z, x)|$$

Triangle inequality $\leftarrow \leq \varepsilon \cdot d_2(x, x')$

- The Laplacian mechanism satisfies the geo-ind condition.

Laplacian mechanism



Optimal mechanisms for geo-ind

- Minimize quality loss (i.e., $E\{d_q(X, Z)\}$) subject to ε -geo-ind constraint.
- Fact: ε -geo-ind constraint is kept under any adversarial remapping $r(\hat{X} | Z)$
- Optimal mechanism is then

$$f^*(Z | X) = \arg \min E\{d_q(Z, X)\}$$

where

$$E\{d_q(X, Z)\} = \sum_{X, Z} f(Z | X) \pi(X) d_q(X, Z)$$

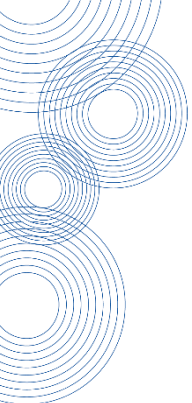
- The optimal adversarial remapping would find

$$r^*(\hat{X} | Z) = \arg \min E\{d_p(\hat{X}, X) | Z\}$$



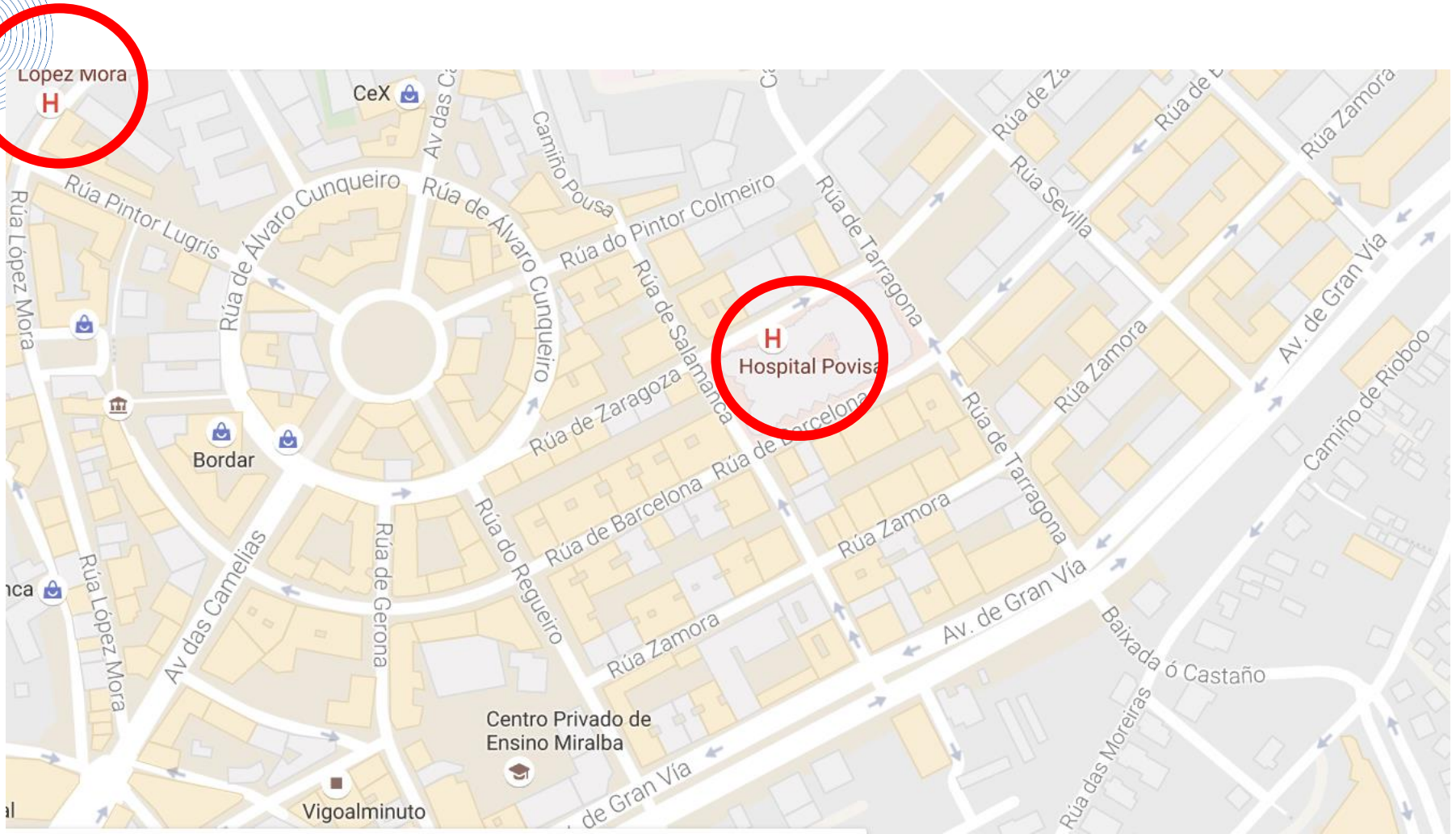
Optimal mechanisms for geo-ind

- If $d_p = d_q$ the adversary does nothing. Minimization of the QL has been already done by the mechanism!!
- But if the adversary does nothing, Privacy=QL.
- The operating value thus depends on ε (the smaller, the larger the privacy).



Where are we going?

Sensitivity [Bertino et. al 2010]



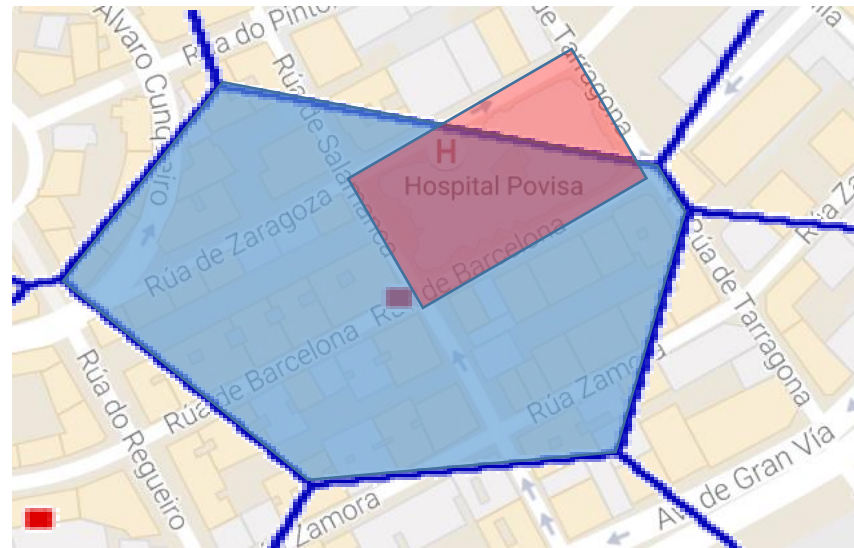
Sensitivity

- The mechanism should weigh the importance given by the user to each location.
- This can be specified semantically by defining categories.

- **Sensitivity of a region:**

prob. that the user, known to be in that region, is actually in a sensitive place.

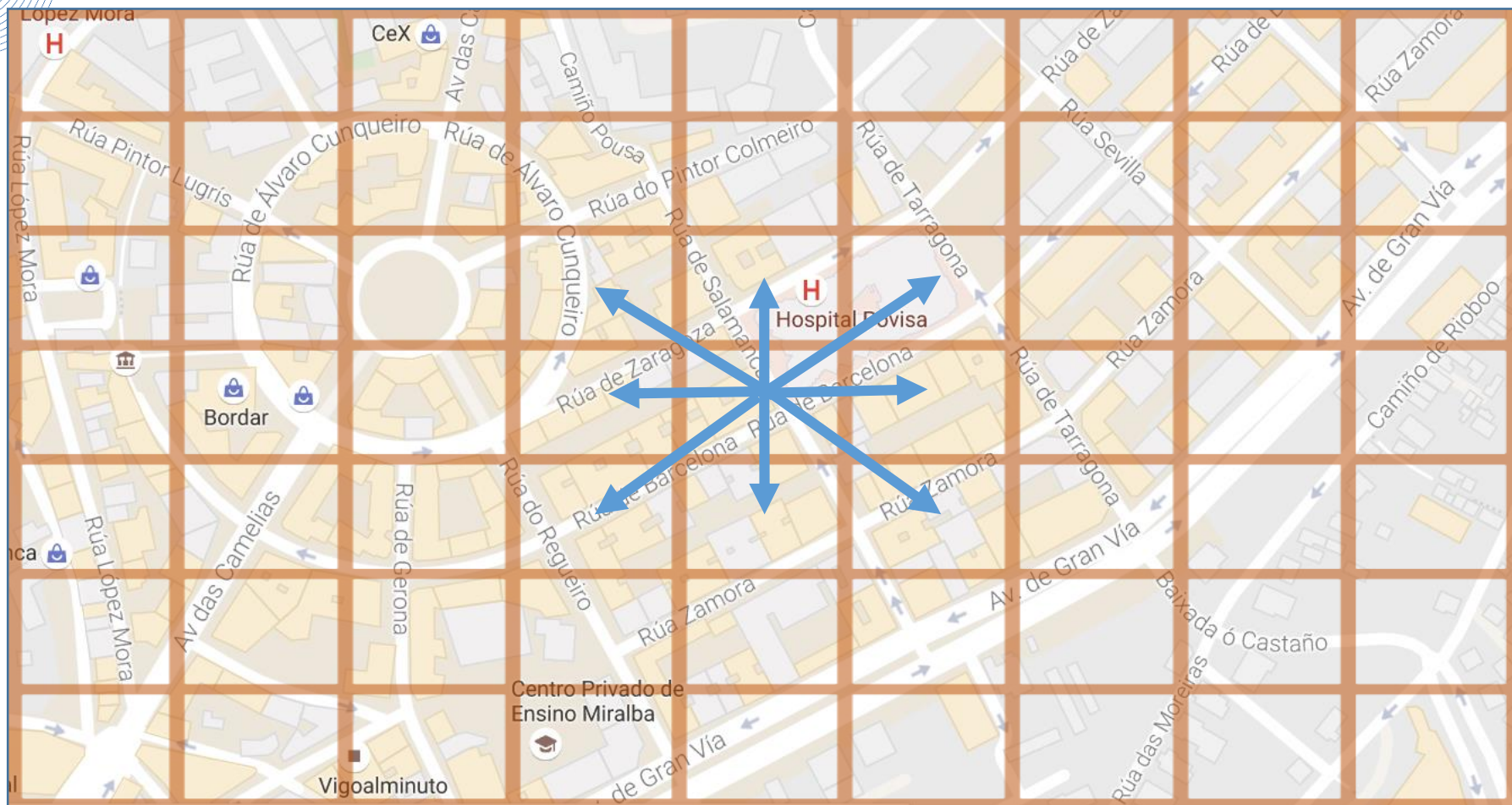
- For other mechanisms: open problem.



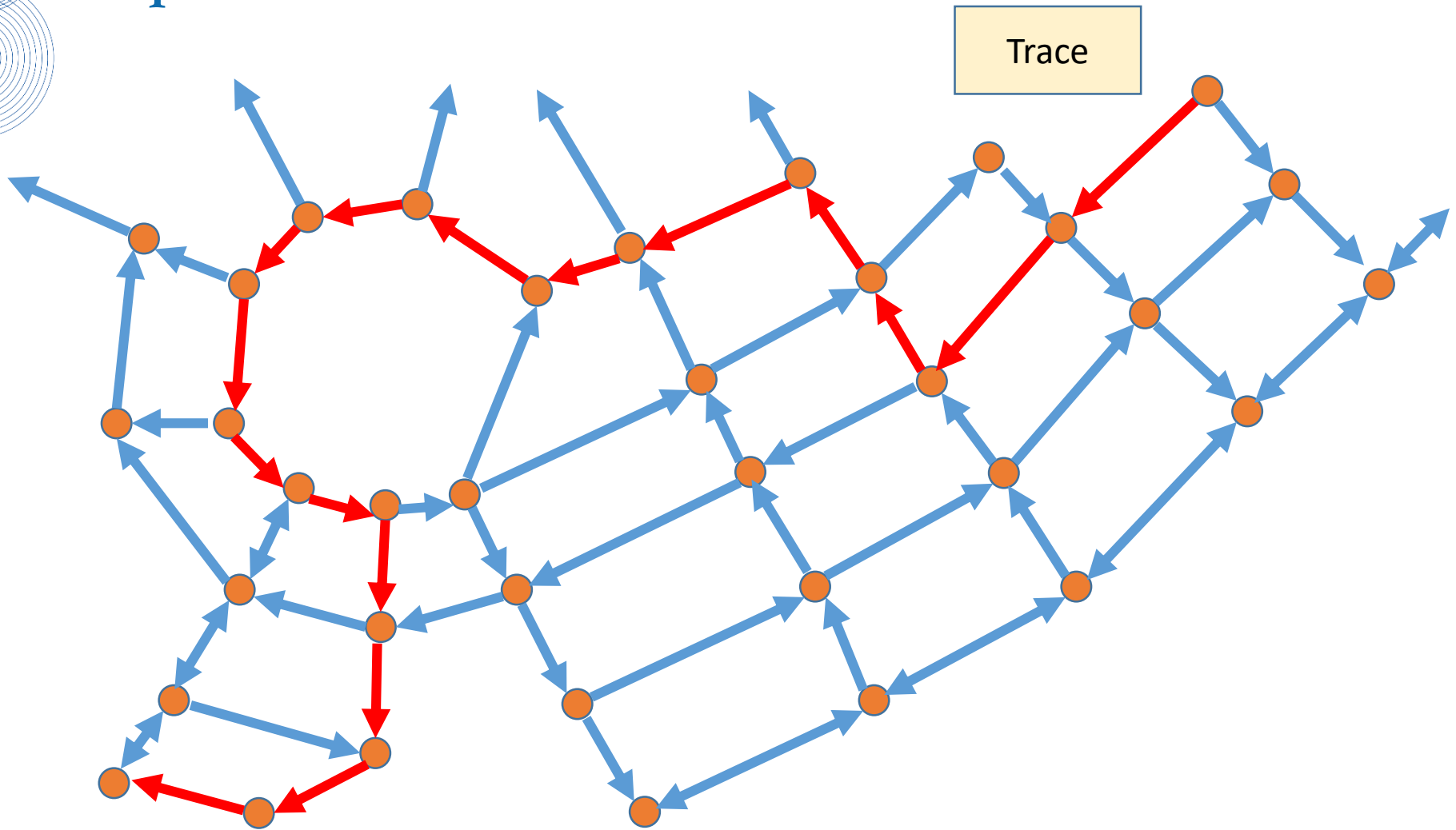
Graph-based models



Graph-based models

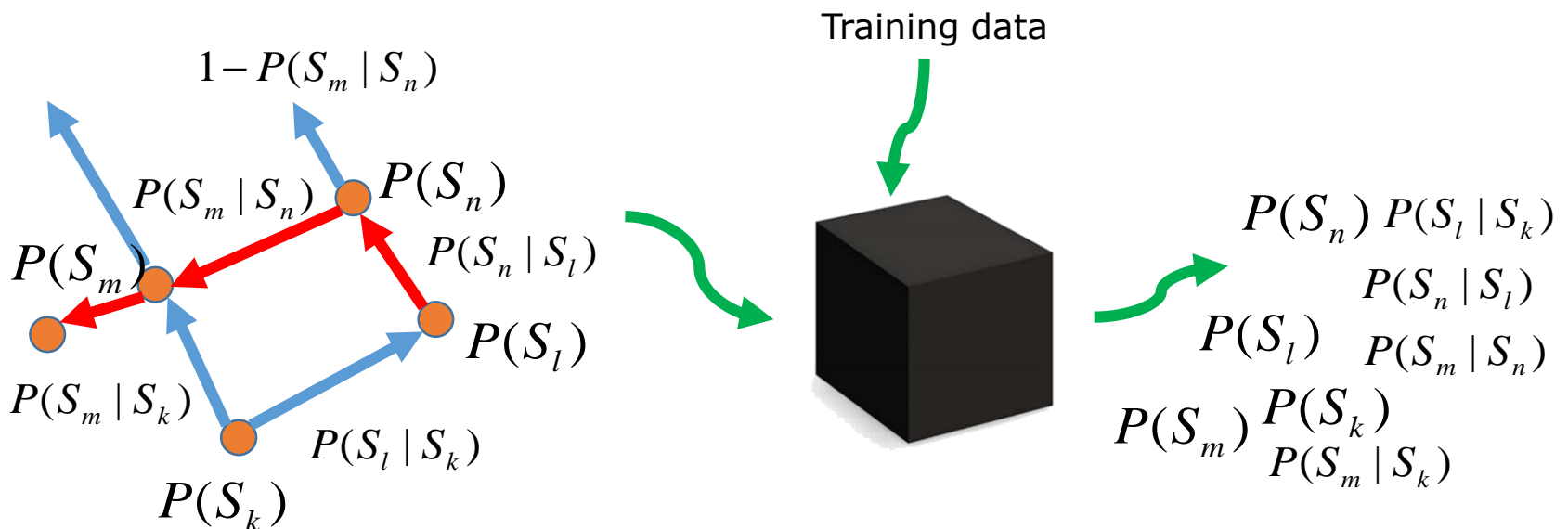


Graph-based models



Graph-based models

- A trace is a path together with time $\{X_i, t_i\}_{i=1}^N$.
- Common assumption for an adversary: the true trace can be described through a Markov chain.
- Prior transition probabilities between states can be estimated if training traces are (at least partially) available.

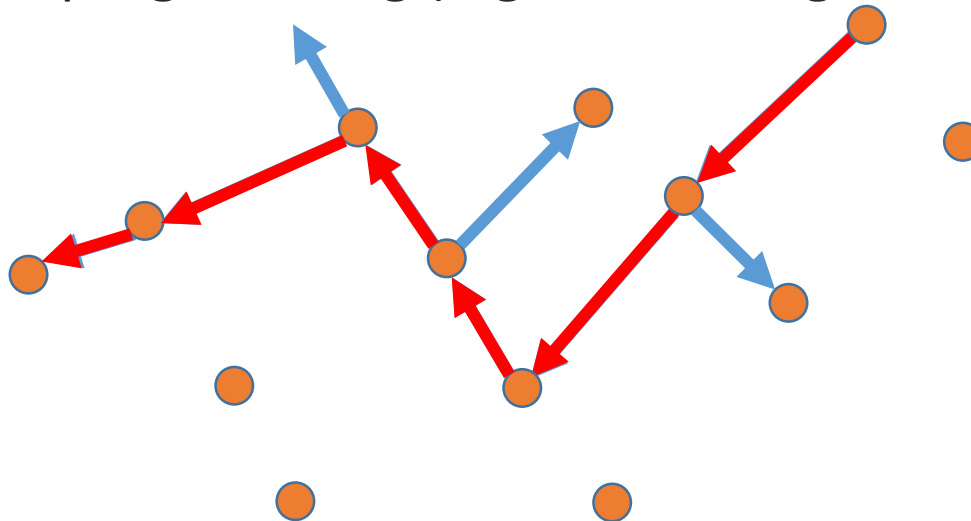


Graph-based models

- Shokri et al.'s approach: depending on what the adversary wants to learn, apply a different method.
- **Maximum likelihood**: find the most likely trace given the observed trace

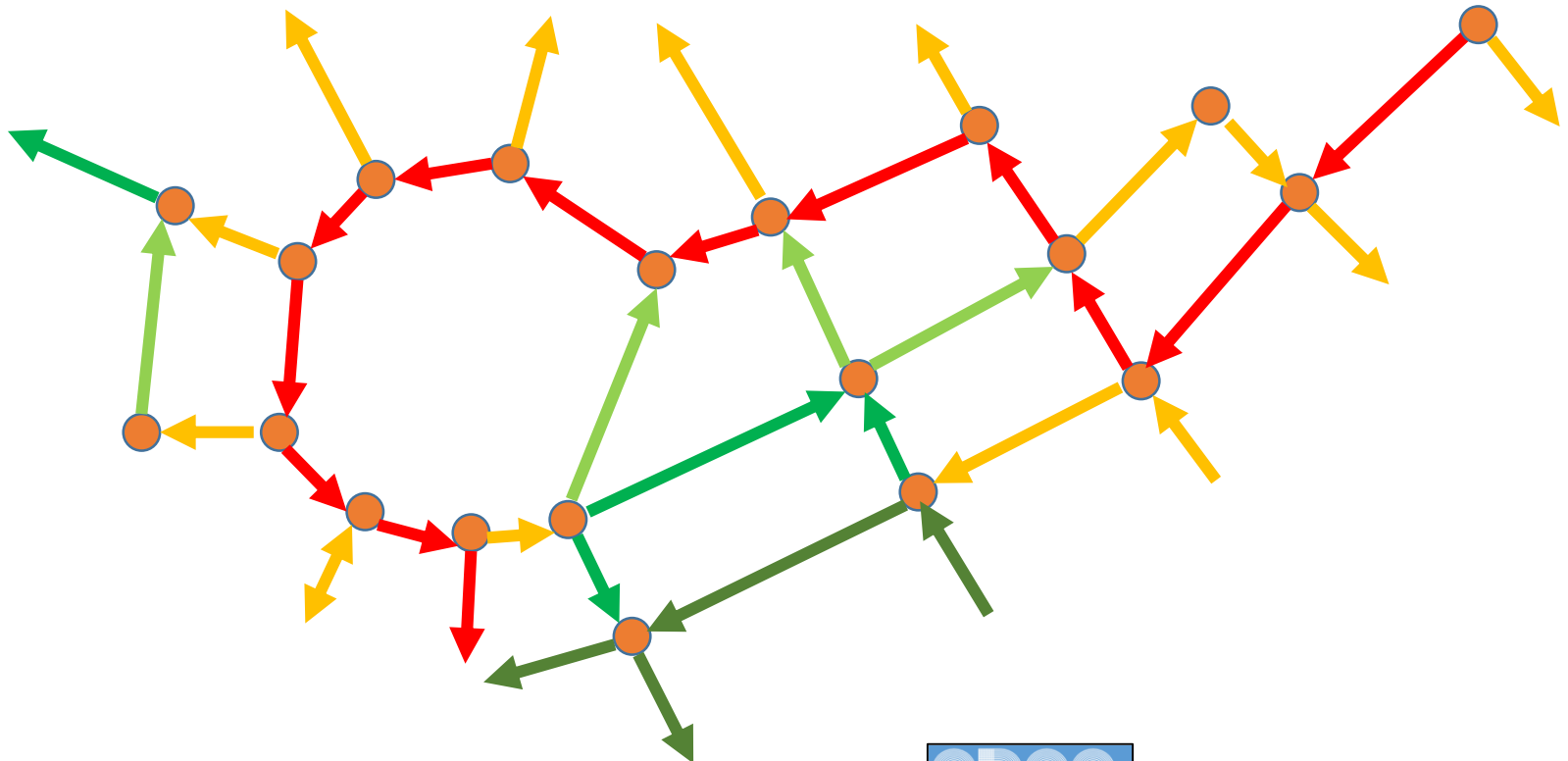
$$\arg \max_{\{X_i, t_i\}_{i=1}^N} f(\{X_i, t_i\}_{i=1}^N \mid \{Z_i, t_i\}_{i=1}^N)$$

- Dynamic programming (e.g., Viterbi algorithm) can be used.



Graph-based models

- **Distribution estimation:** estimate the probabilities of all traces using the Metropolis-Hastings algorithm.

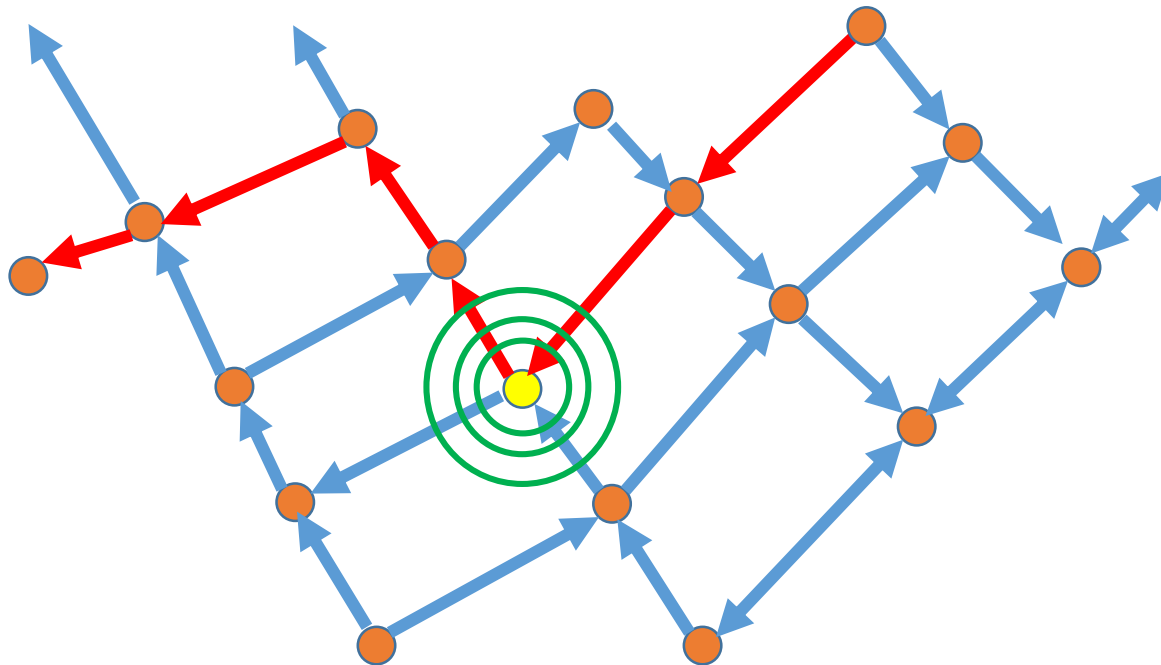


Graph-based models

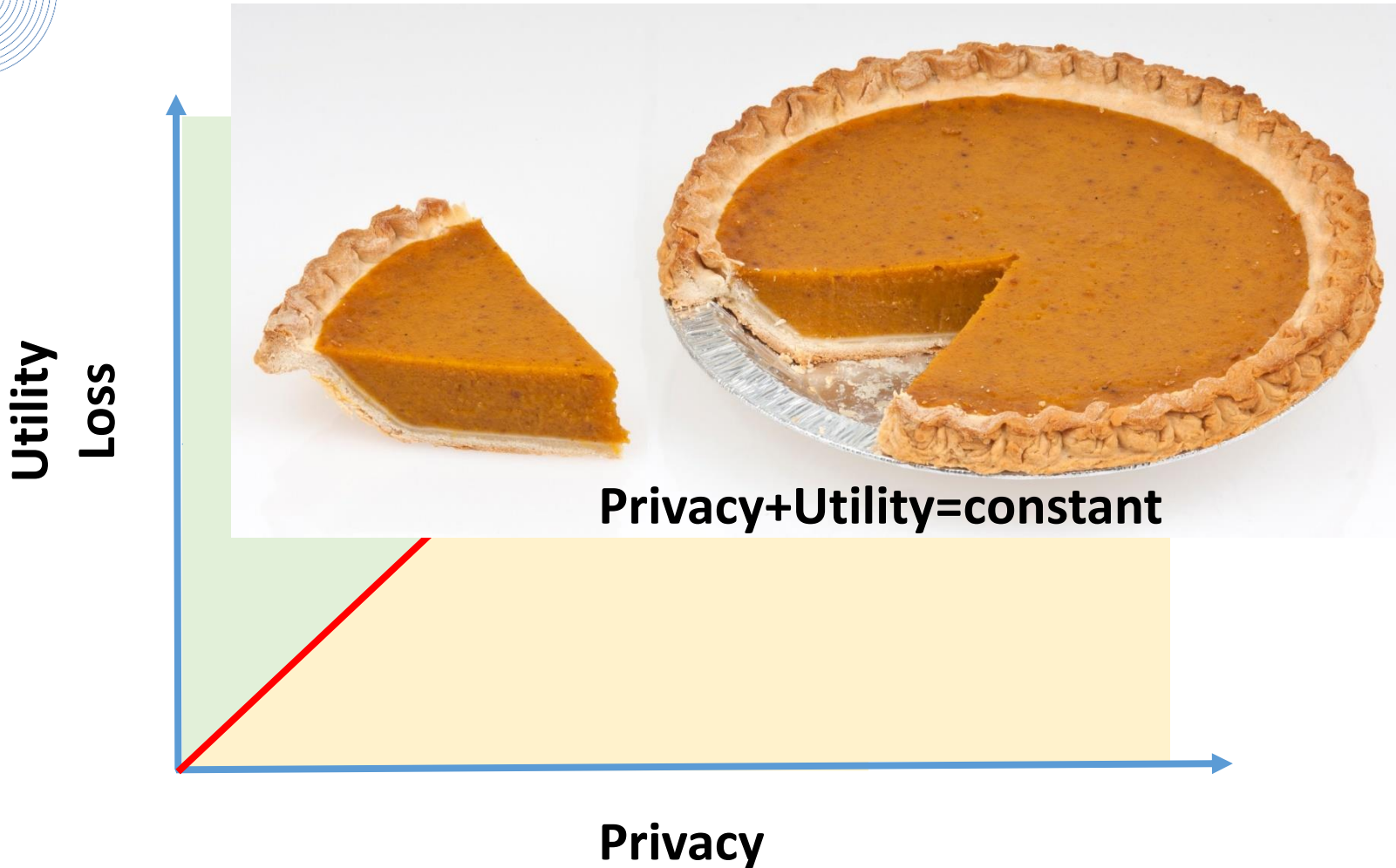
- **Location estimation:** find the most likely node at time t_k

$$\arg \max_{X_k} f(X_k | \{Z_i, t_i\}_{i=1}^N)$$

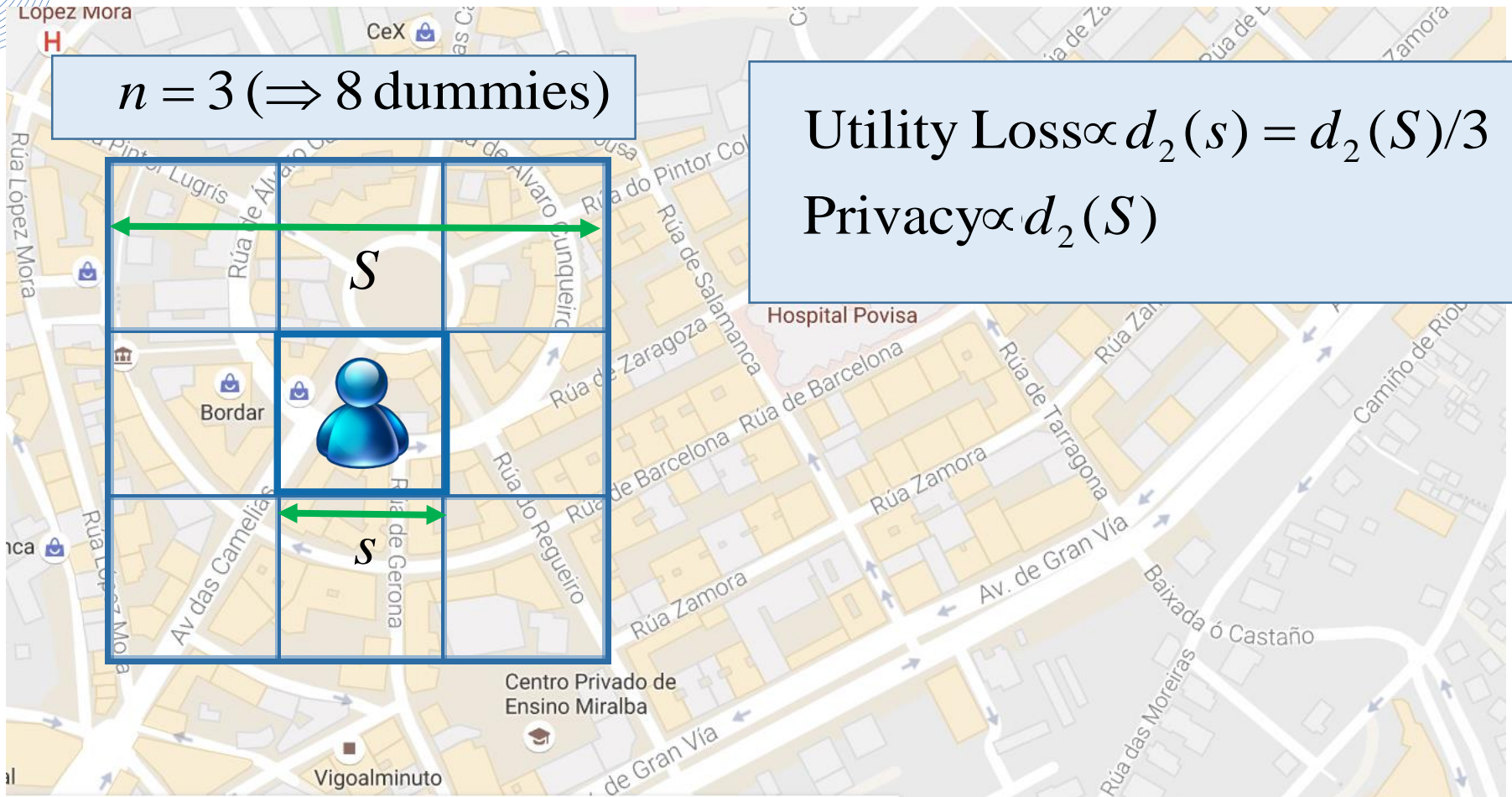
- Can be solved using the backward-forward algorithm to recursively compute the probabilities.



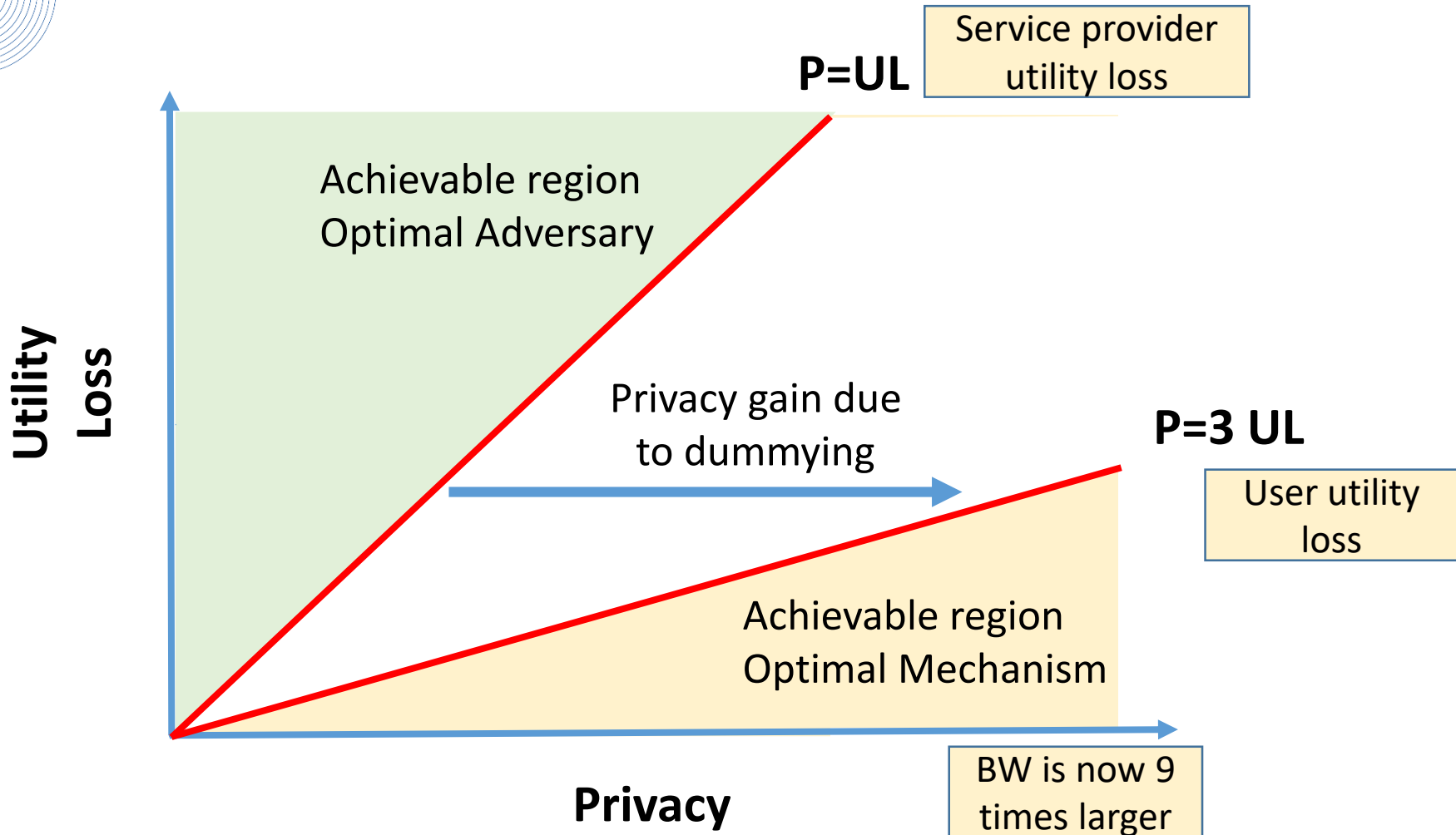
Privacy as a zero-sum game



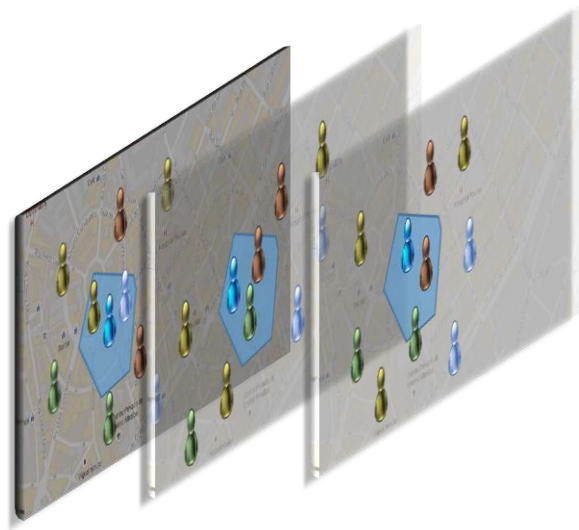
Adding a new dimension: bandwidth



The Utility Loss-Privacy-Bandwidth region



Space-time cloaking



Utility Loss \propto area

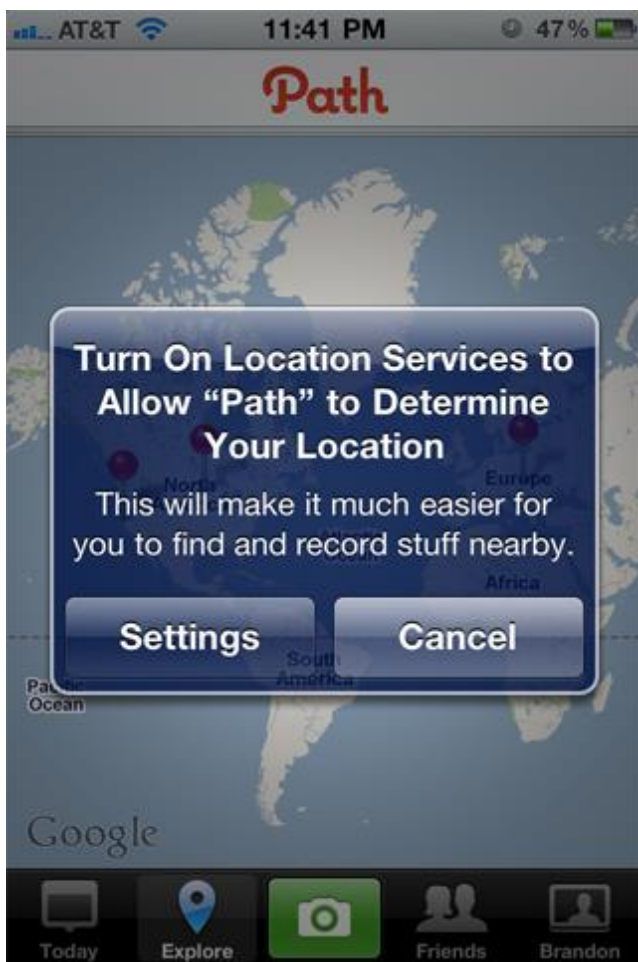
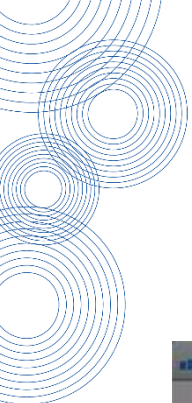
Privacy \equiv k - anonymity \propto area \times time \times pop. density

Delay \propto time

Privacy-preserving queries

Retrieval in Encrypted Domain





AtlantTIC

Research Center for
Information & Communication Technologies



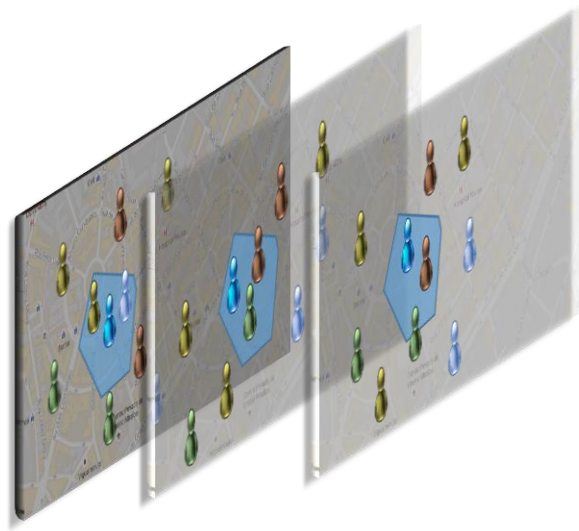
Grupo Procesado de
Señal en Comunicaciones

Thanks!

fperez@gts.uvigo.es

www.gpsc.uvigo.es

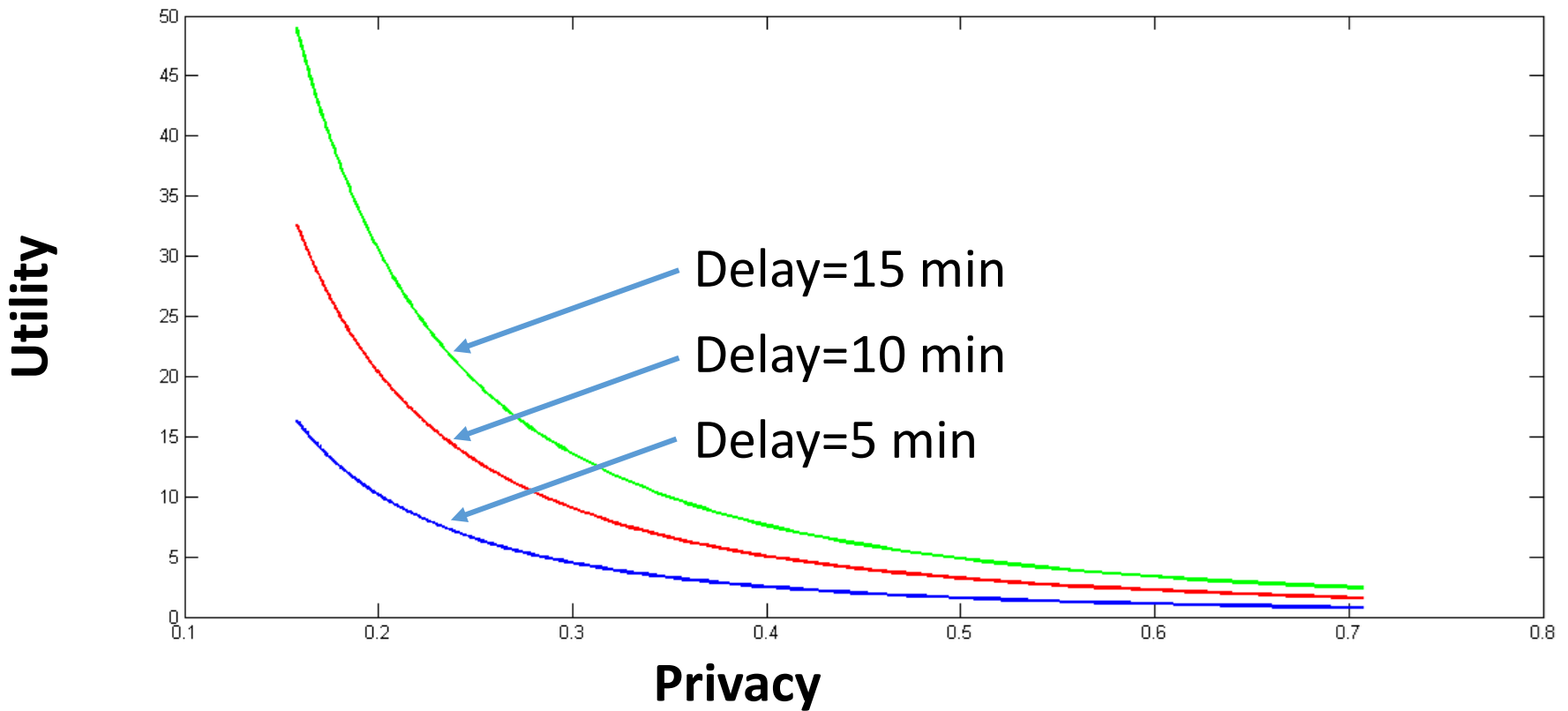
What utility? An example

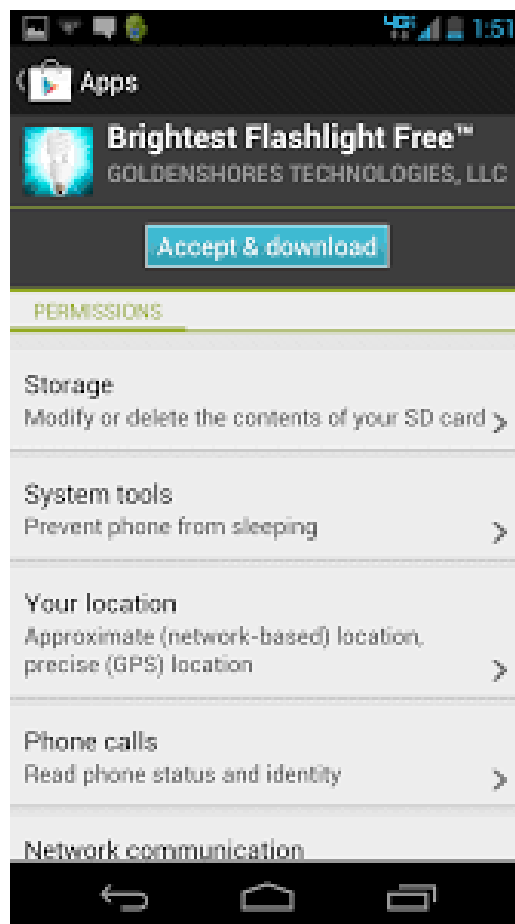
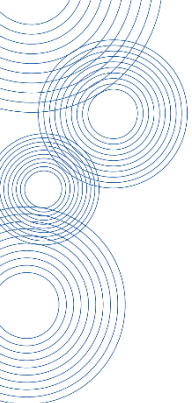


$$\text{Utility} \equiv 1 / d_{\max} \propto 1 / \sqrt{\text{area}}$$

$$\text{Privacy} \equiv k\text{-anonymity} \propto \text{area} \times \text{time} \times \text{pop. density}$$

But delay also counts...

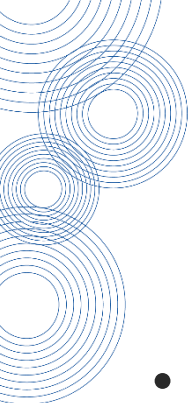






What utility? Another example

- Space-time slicing
- Is this related to bandwidth?



- Space-time slicing
- Is this related to bandwidth?