CS-523 Advanced Topics on Privacy Enhancing Technologies

Location privacy

Theresa Stadler
SPRING Lab
theresa.stadler@epfl.ch

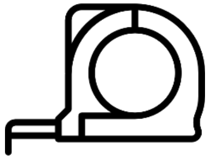
- Some slides/ideas adapted from: Carmela Troncoso, George Danezis, Jean-Pierre Hubaux, Reza Shokri

Introduction
Location privacy

Course aim: learn **toolbox for privacy engineering**

toolbox
around location privacy implications

notions
to express location privacy



tools
to quantify location privacy

tools
to mitigate location-related inferences

Application Layer

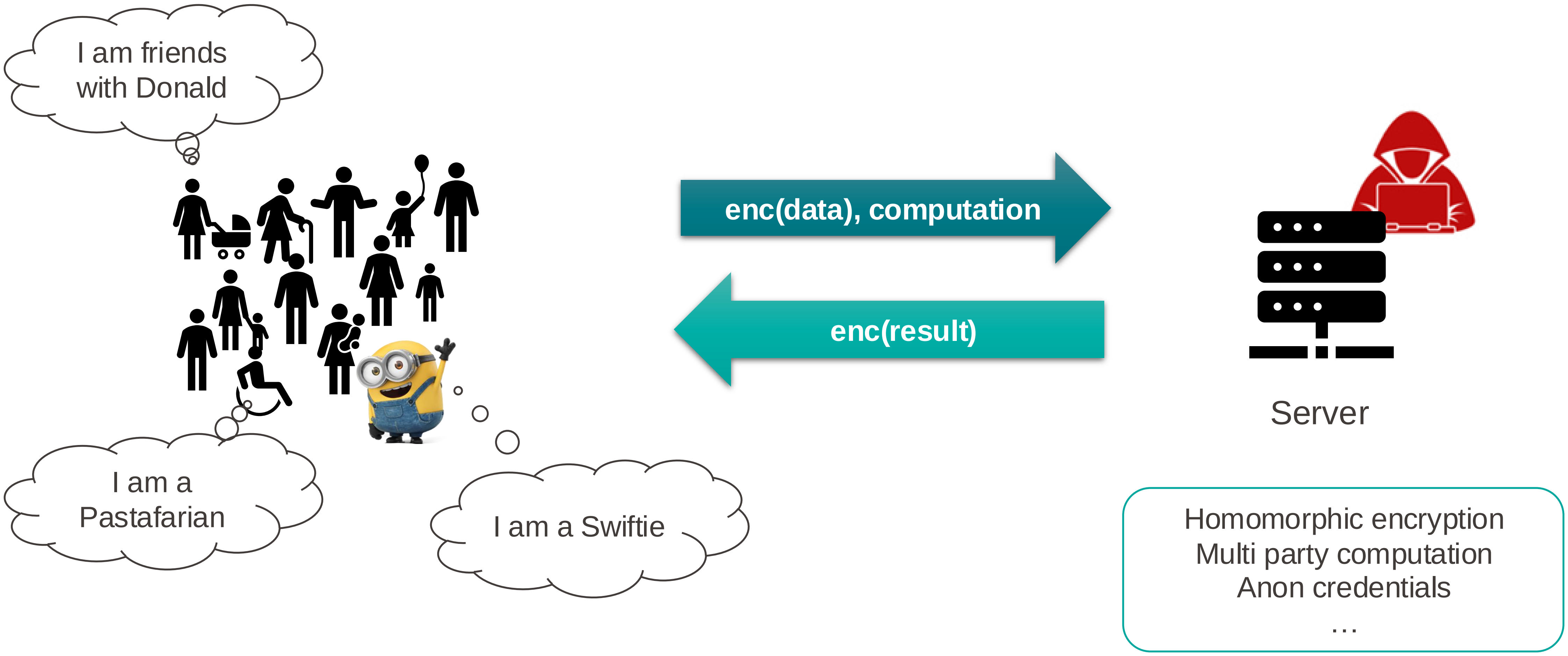
Network Layer

Goals

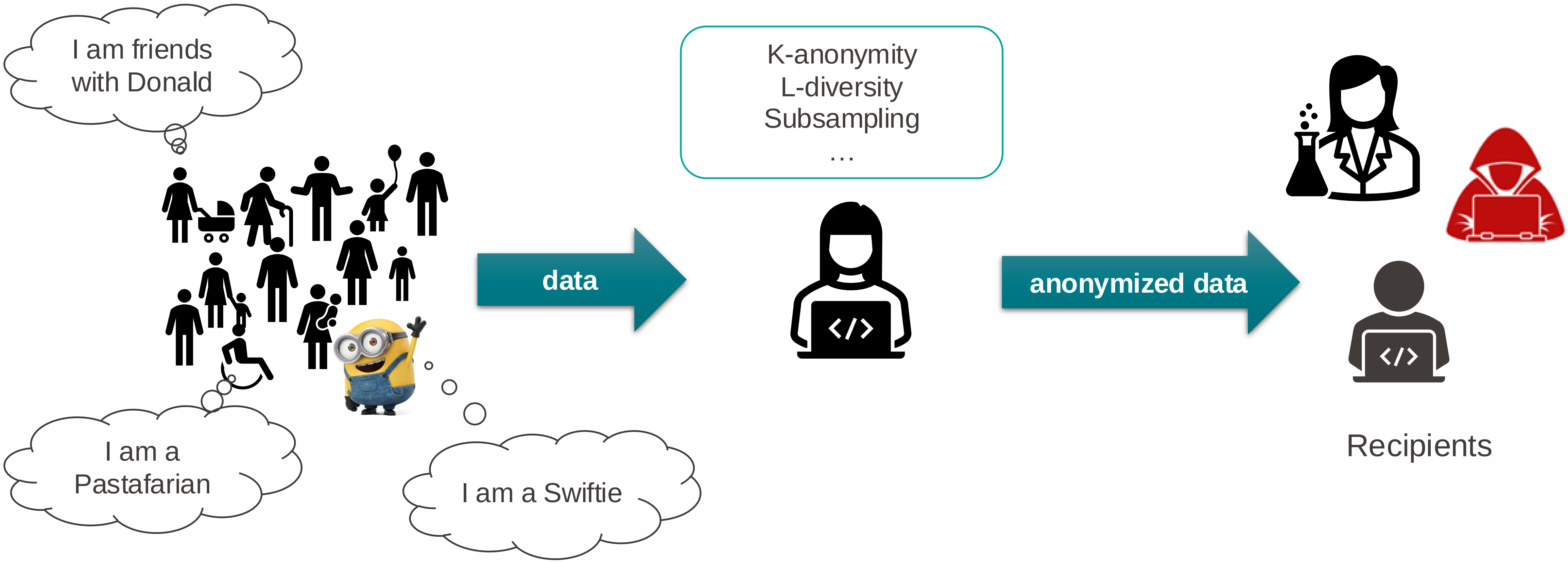
What should you learn today?

- Understand protecting privacy requires **more than hiding contents**
- Understand the **privacy issues of location data**
  - Trust assumptions
  - **Adversarial models**
- Understand how to protect **location privacy**
  - **How to mitigate** adversarial inference capability
  - **How to quantify** privacy loss
- Understand **practical issues** when protecting individuals' whereabouts
  - It is very, very, very hard (no known way to get good protection)

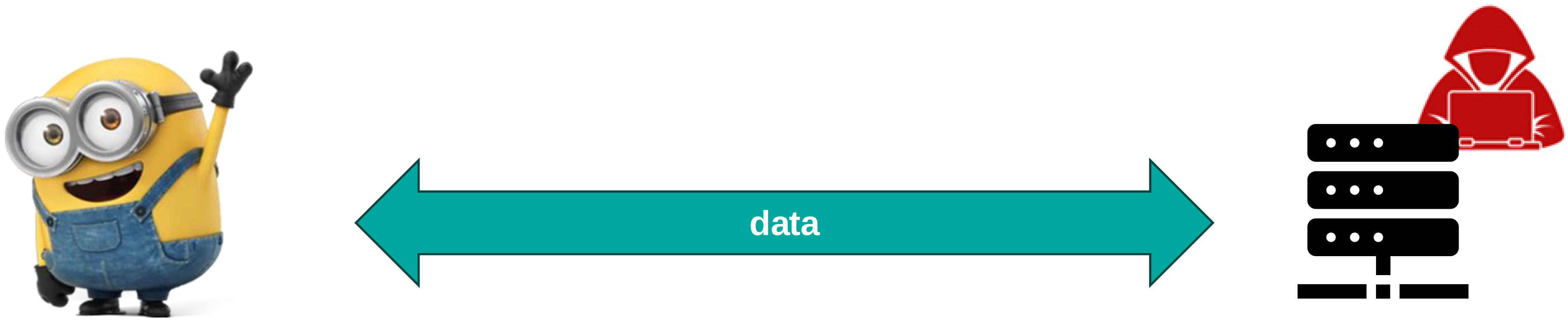
Common thought: Privacy is all about data



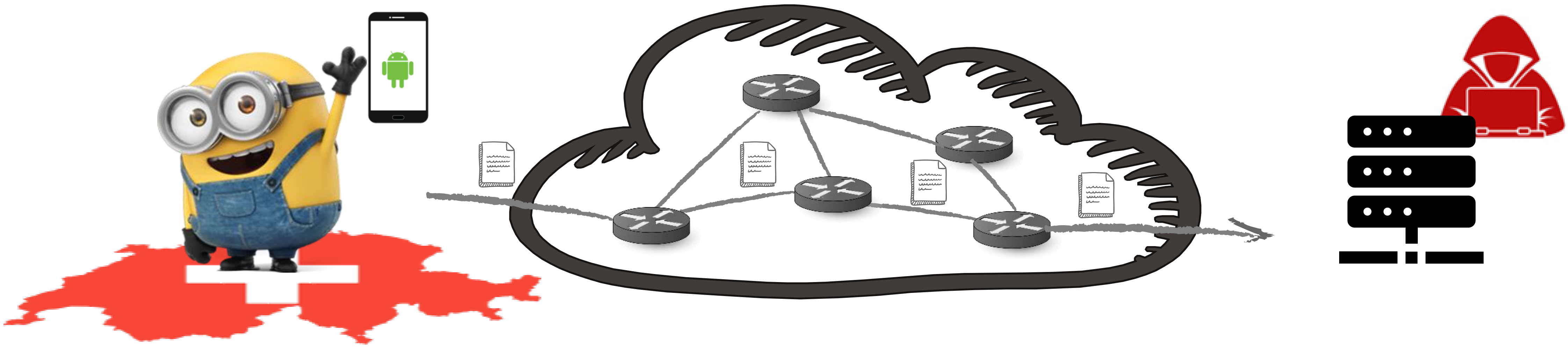
Common thought:
Privacy is all about data



This might be a good model if the world was like this…



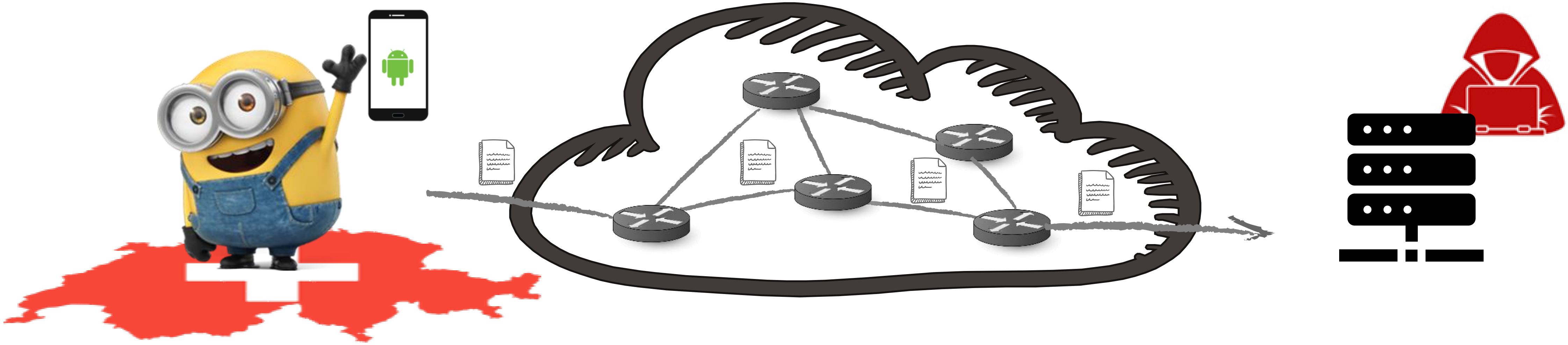
But in reality…



But in reality…

Device type, OS, applications/software, sensors,….

IP, MAC, routes,…



Location

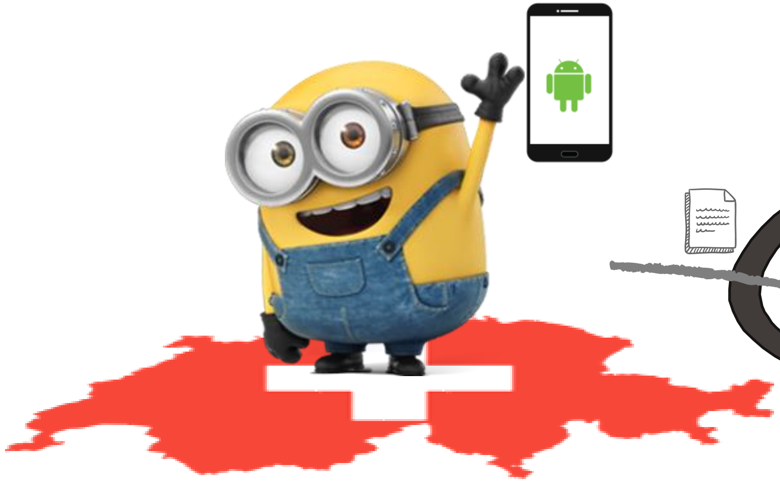
Metadata all around

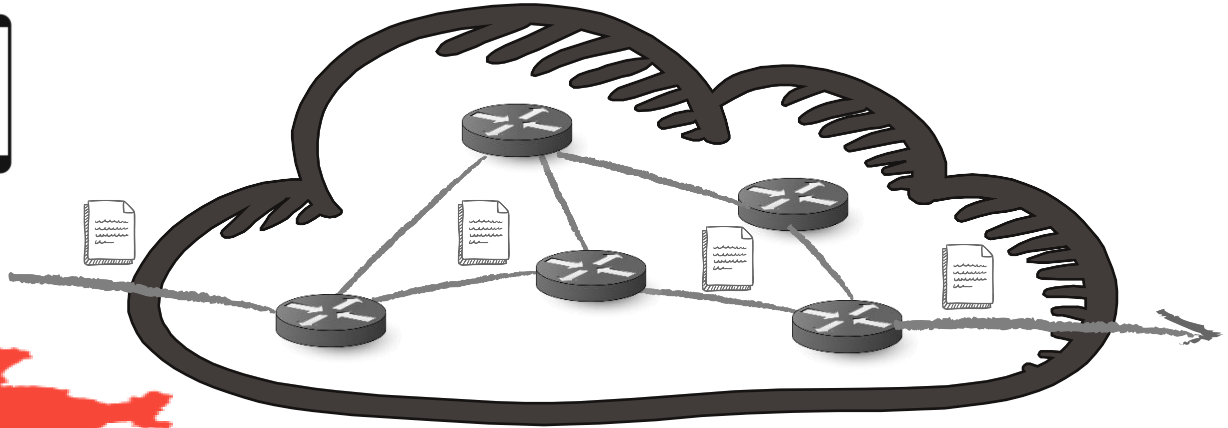
Metadata encodes a lot of information

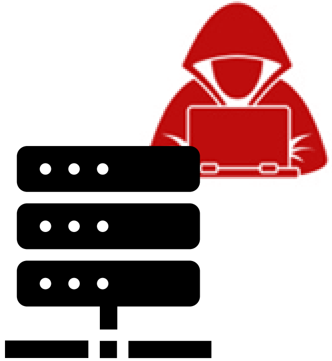
Pseudoidentifier | Device type, OS,

Pseudoidentifier
Sensitive | applications/software
, sensors,….

Pseudoidentifier
Location | IP, MAC, routes,…







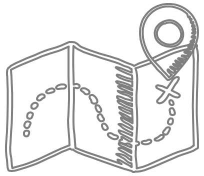
Pseudoidentifier
Sensitive | Location

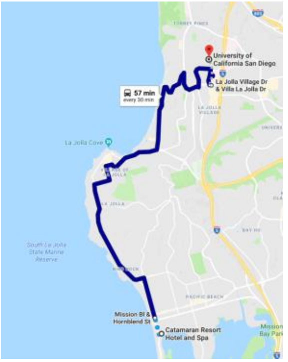
Metadata all around

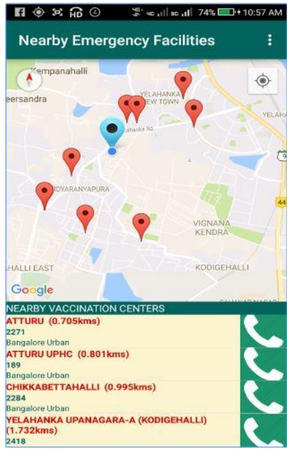
Location privacy

Location data is useful…

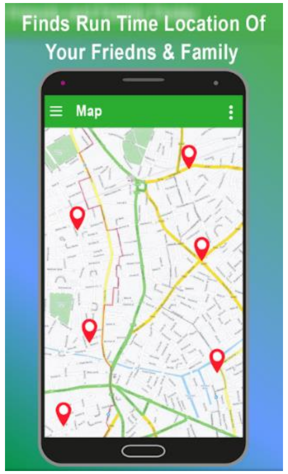
Bob

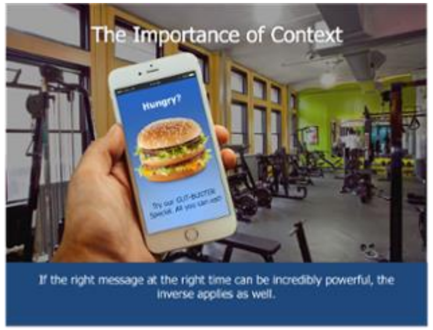






Location Intelligence

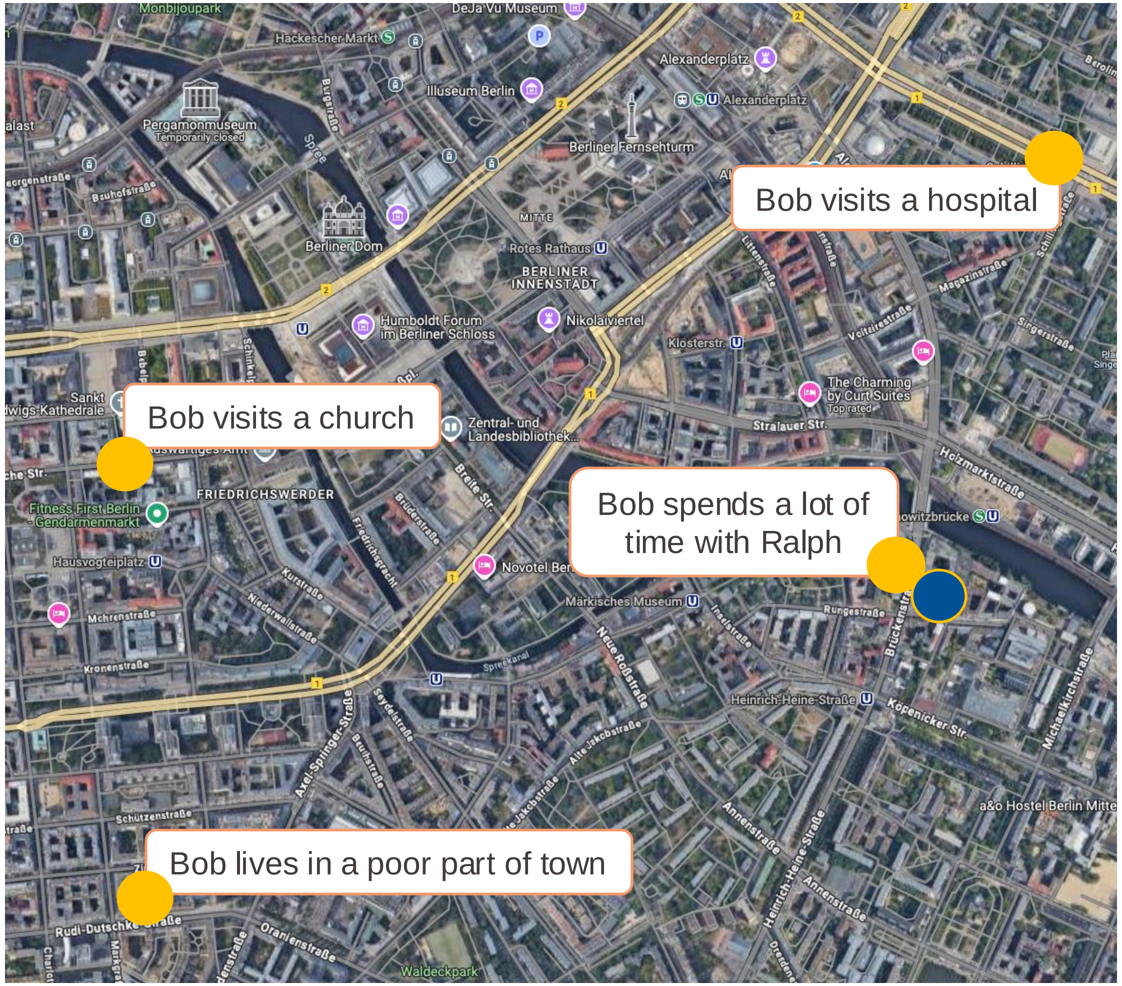




Google search hits
1770M in Mar'24
(749M in Apr'22)
(653M in Apr'20)
(355M in Apr'19)
(197M in Jul' 18)
(649K in Feb' 18)

But contains a lot of sensitive information

About our religious beliefs



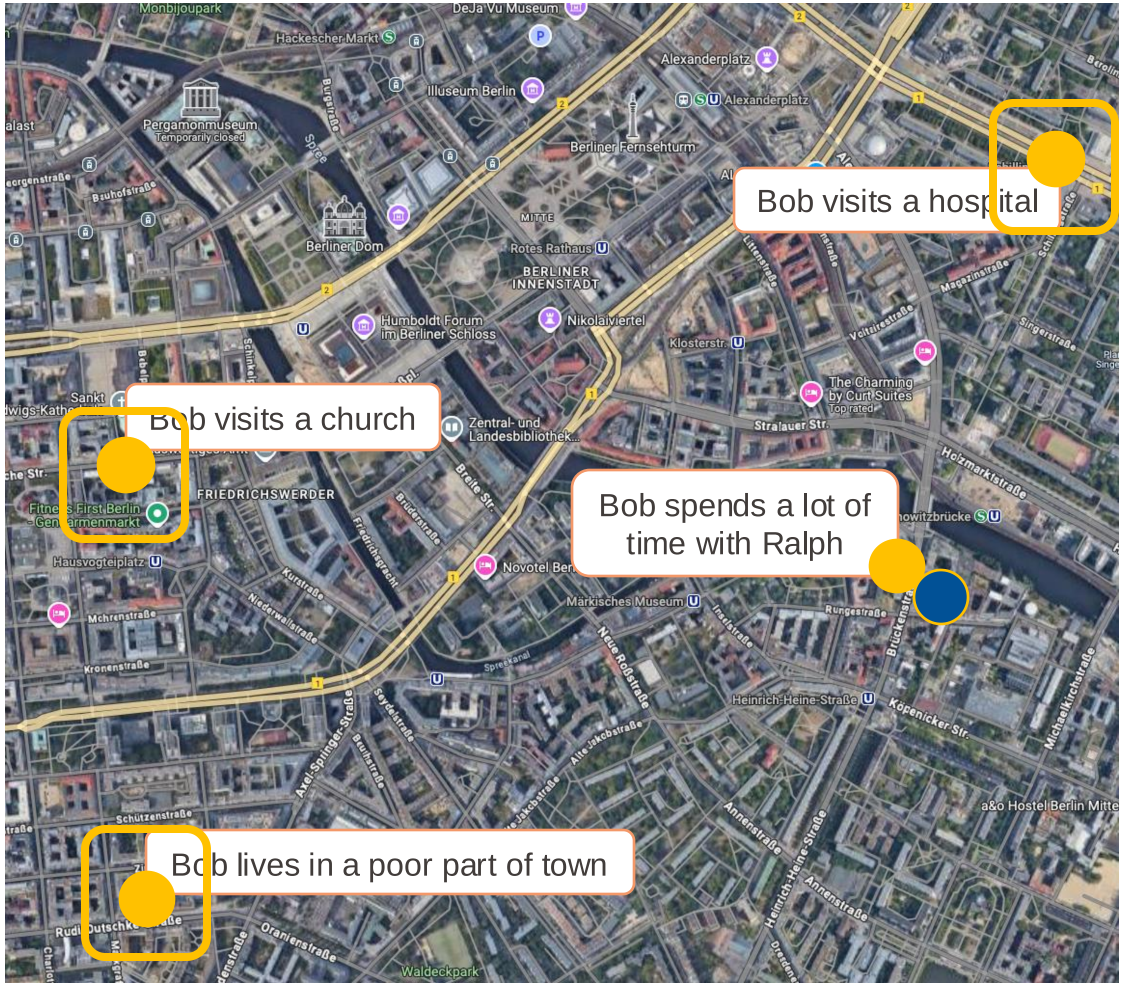
About our health status

About our financial situation

About our social relationships

Inference: Points of interest (POIs)

What is a POI? A specific location that someone may find useful or interesting



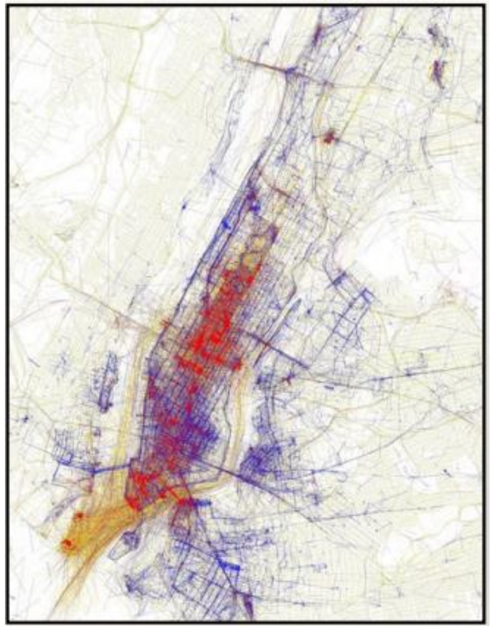
Inference: Points of interest (POIs)

Why are POIs important? Because our movements are unique

[De Montjoye et al 2013] [De Montjoye et al 2015]: *4 spatio-temporal points are enough to uniquely identify 95% of people in a mobile phone database of 1.5M people and to identify 90% of people in a credit card database of 1M people*

Four known points you were at

+



Anonymized location dataset

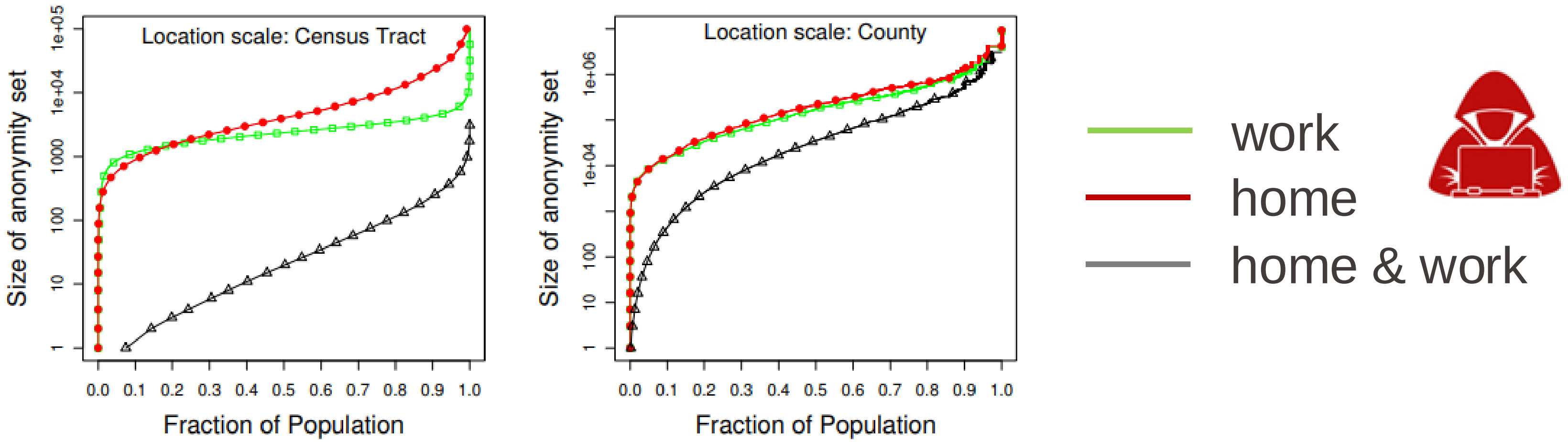
=

All your whereabouts

Inference: Points of interest (POIs)

Why are POIs important? Because our home & work location are unique identifiers

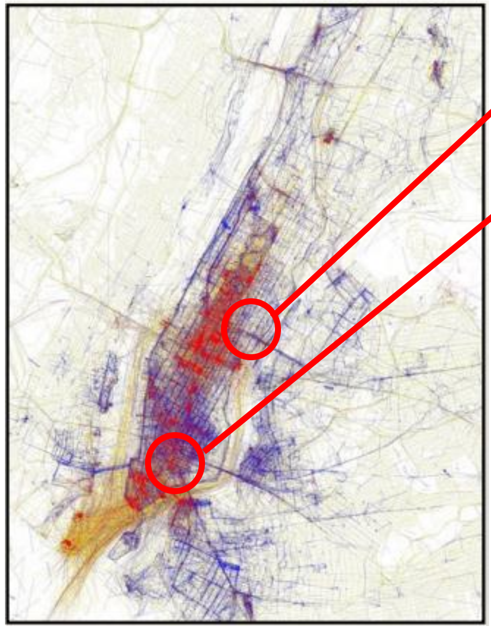
[Golle & Partridge 2009] *given home & work*, median *individual's anonymity set in the U.S. working population is 1, 21 and 34,980, for locations known at the granularity of a census block, census track and county respectively*



Inference: Points of interest (POIs)

Why are POIs important? Because our home & work location are unique identifiers

[Golle & Partridge 2009] *given home & work*, median *individual's anonymity set in the U.S. working population is 1, 21 and 34,980, for locations known at the granularity of a census block, census track and county respectively*

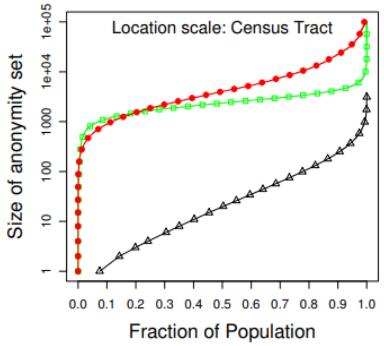


Anonymized location data

Mostly at night

Mostly from 9am to 5pm

\+



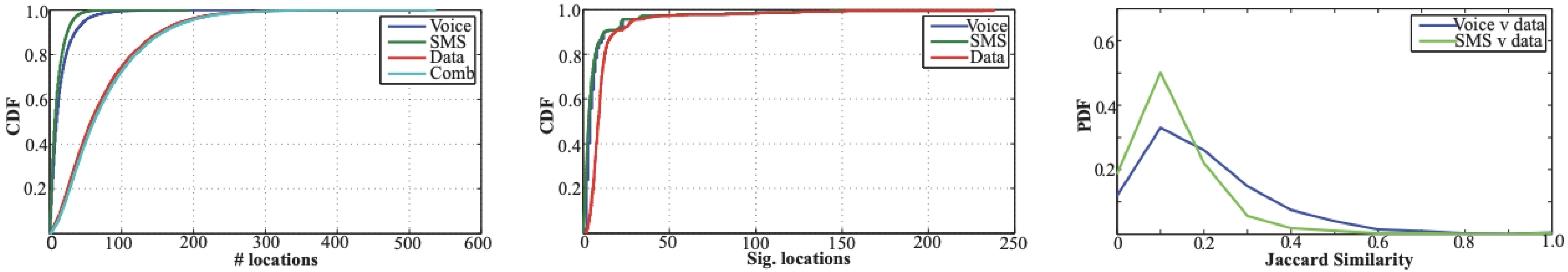
=

All your whereabouts

Inference: Points of interest (POIs)

Why are POIs important? Because our home & work location are easily inferred

[Zhang & Bolot 2011] showed that for voice call and SMS records from cellular networks *“top 2" locations likely correspond to home and work locations, the "top 3" to home, work, and shopping/school/commute path locations*

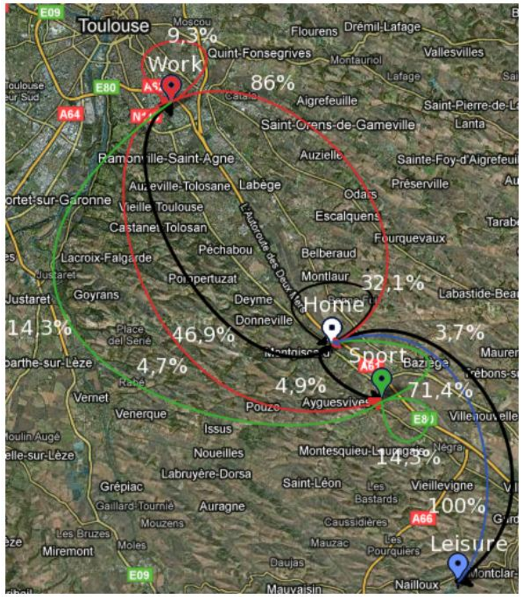


(a) Distinct locations per user (b) Significant loc. (c) Overlap in significant loc.

Inference: Points of interest (POIs)

Why are POIs important? Because they allow to predict where someone moves next

[Gambs et al 2012] *Accuracy for the prediction of the next location in the range of 70% to 95%*

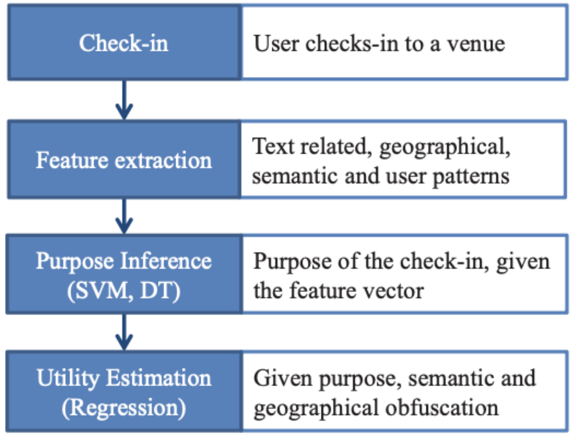


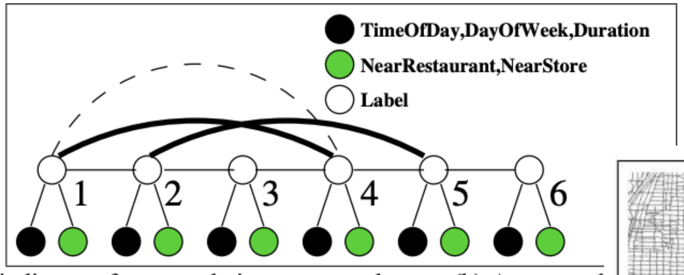
A hidden Markov model of
individual movement patterns

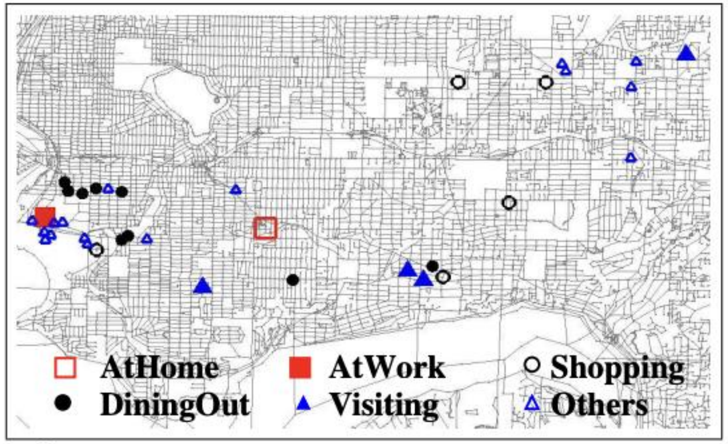
Inference: Points of interest (POIs)

Why are POIs important? Because they allow to infer demographics and other attributes

[Pang and Zhang 2017] [Felbo et al 2017] [Bilogrevic et al 2015] [Cho et al 2010] [Liao et al 2005] [Liao et al 2007] present **machine learning based frameworks** to infer sensitive attributes from location data



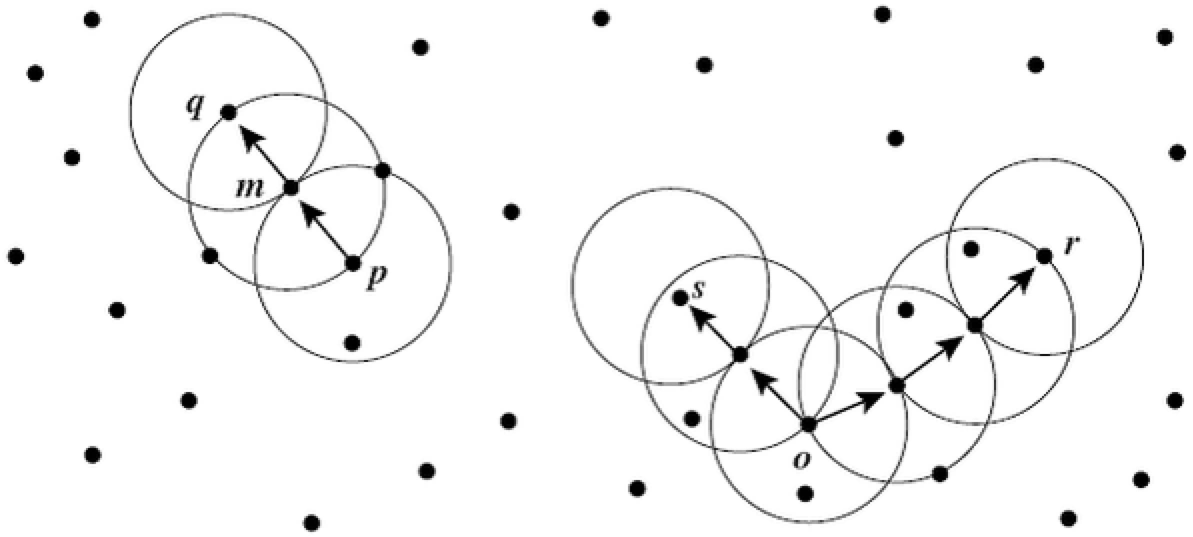




Inference: Points of Interest (POIs)

How to extract POIs? Clustering techniques [Ester et al 1996][Ashbrook & Starner 2003][Krumm 2007]

[Ester et al 1996]: Simple yet effective way to infer POIs: **DBSCAN**



And after finding the clusters/POIs?

Home and work: identified by time

Further split clusters
(e.g., using X-Means [Pelleg & Moore 2000])

Inferences can be automatized using reverse geo-coding (e.g., on the centroids)!

https://www.datacamp.com/tutorial/dbscan-clustering-algorithm

So where does all of this data come from?

At the application level

- User location revealed as part of application functionality
- Application might access location for personalization (or tracking)
- Location might be revealed trough metadata of files accessible by the application e.g. images

At the network level

- IP-based geolocation
- WiFi access points (SSIDs, MAC addresses)
- Bluetooth beacons

Likely many more…

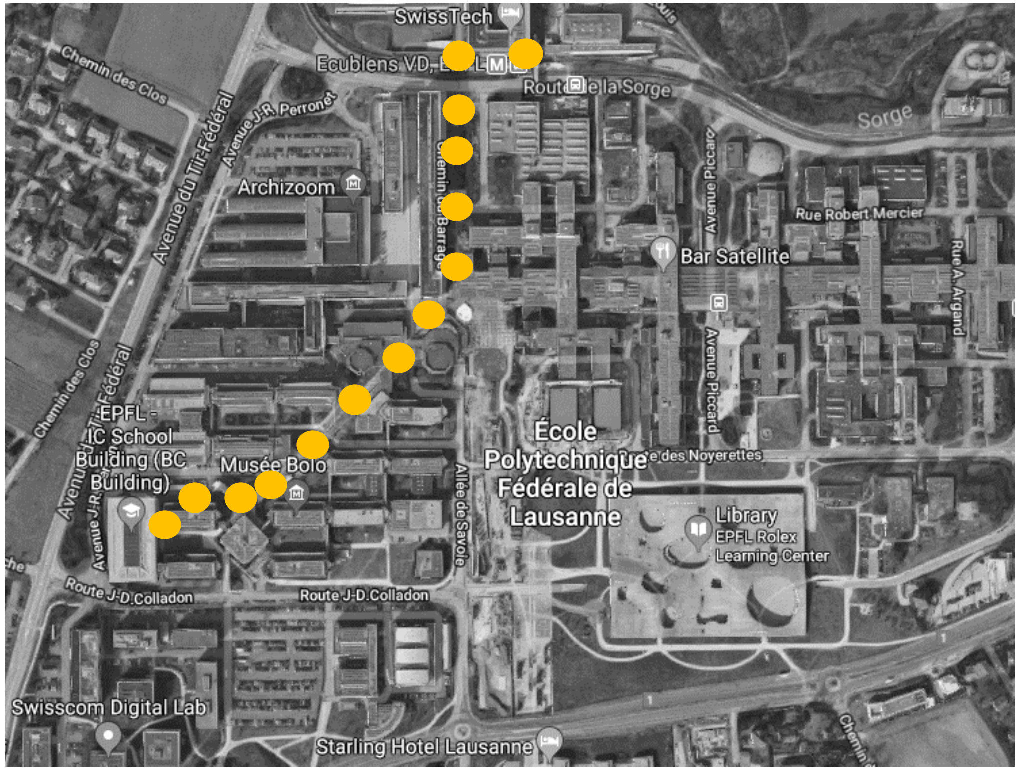
EPFL

How to protect location privacy

How to protect location privacy

4 main techniques:

Perturbation
Hiding
Generalization
Adding dummies



How to protect location privacy
Perturbation

Spatial Obfuscation: Perturbation of locations using noise [Duckham & Kulik 2005]

Geo-indistinguishability [Andres et al. 2013]

If two points are close, their obfuscated points are close

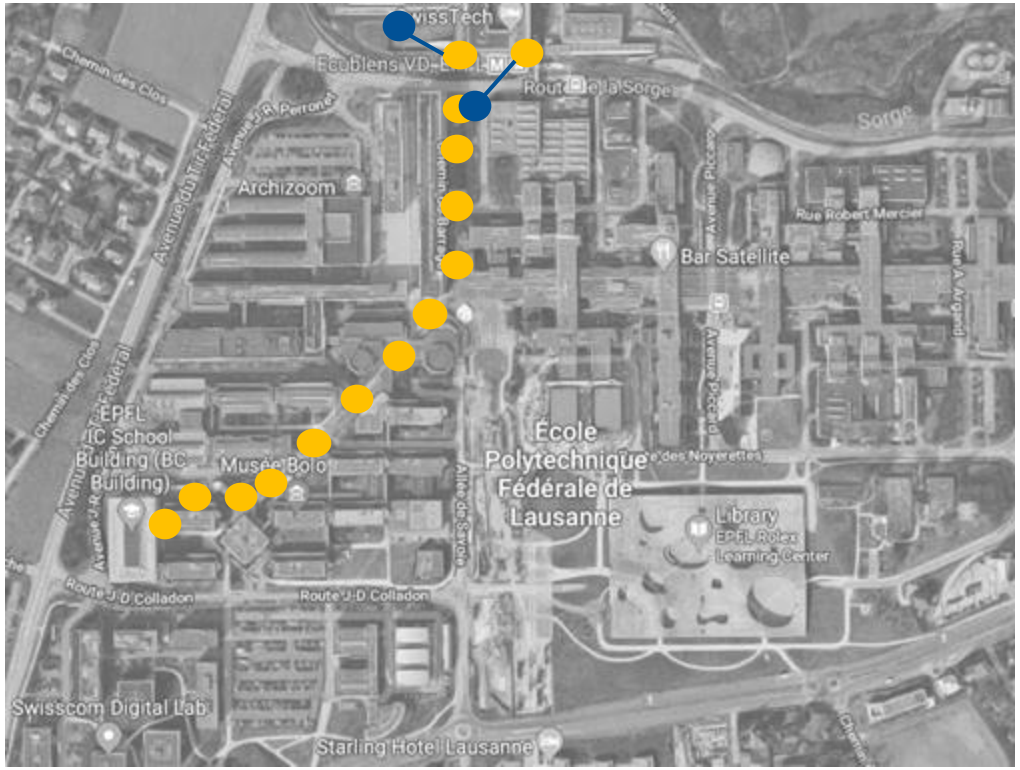
obfuscation mechanism

distance measure

$$d[{\color{blue}\bullet} = K({\color{orange}\bullet}) \mid {\color{blue}\bullet}' = K({\color{orange}\bullet}')] < \epsilon d[{\color{orange}\bullet} \mid {\color{orange}\bullet}']$$

obfuscated point

original point



How to protect location privacy
Perturbation

Spatial Obf

Geo-indistir

If two points are

obfuscation me

$d[$● $= K($

obfuscated poin

Differential Privacy
Formal Definition

A mechanism M is ε-differentially private if for all neighbouring databases D and D_{-r} which differ in only one individual

$$\mathbb{P}[M(D) = O] \leq e^{\varepsilon} \cdot \mathbb{P}[M(D_{-r}) = O]$$

… and this must be true for all possible outputs O

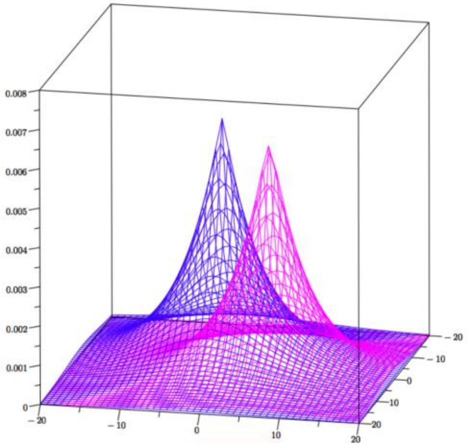
How to protect location privacy
Perturbation

Spatial Obfuscation: Perturbation of locations using noise [Duckham & Kulik 2005]

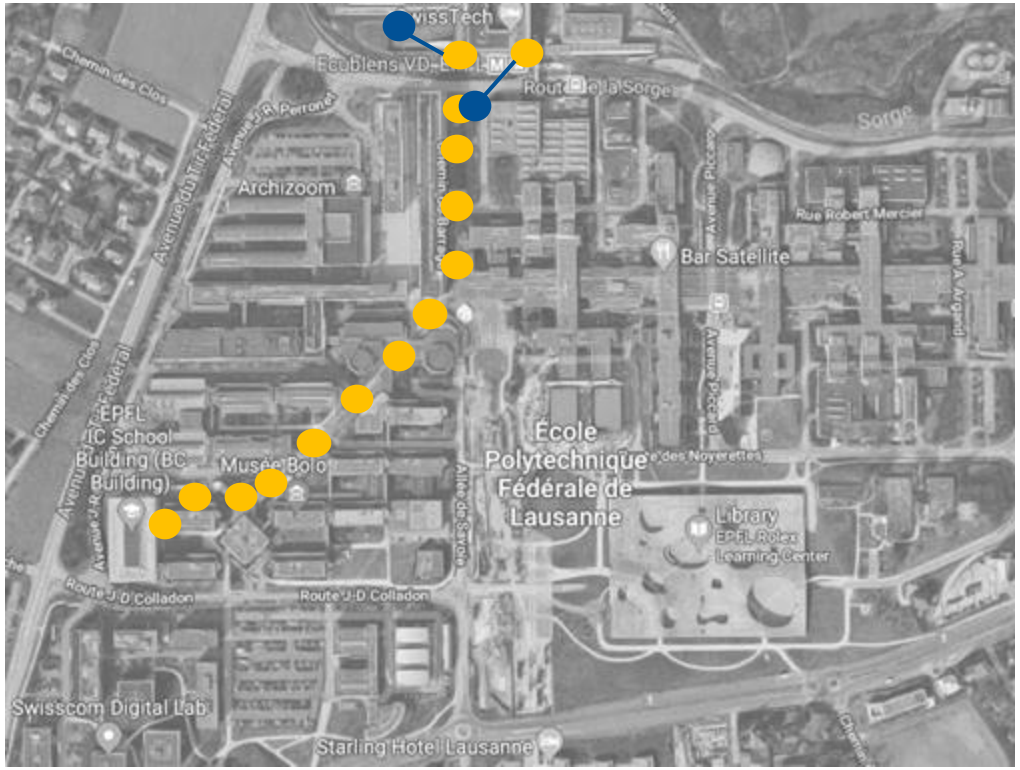
Geo-indistinguishability [Andres et al. 2013]

If two points are close, their obfuscated points are close

$$d[\bullet = K(\bullet) \mid \bullet' = K(\bullet')] < \epsilon d[\bullet \mid \bullet']$$



Add 2-dimensional
ϵ-differential privacy noise



How to protect location privacy
Perturbation

Spatial Obfuscation: Perturbation of locations using noise [Duckham & Kulik 2005]

Geo-indistinguishability [Andres et al. 2013]

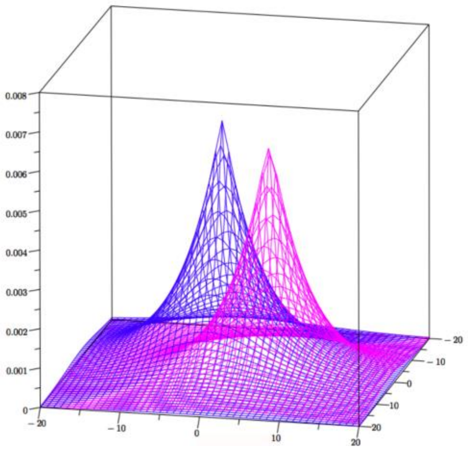
If two points are close, their obfuscated points are close

$$d[\bullet = K(\bullet) \mid \bullet' = K(\bullet')] < \epsilon d[\bullet \mid \bullet']$$

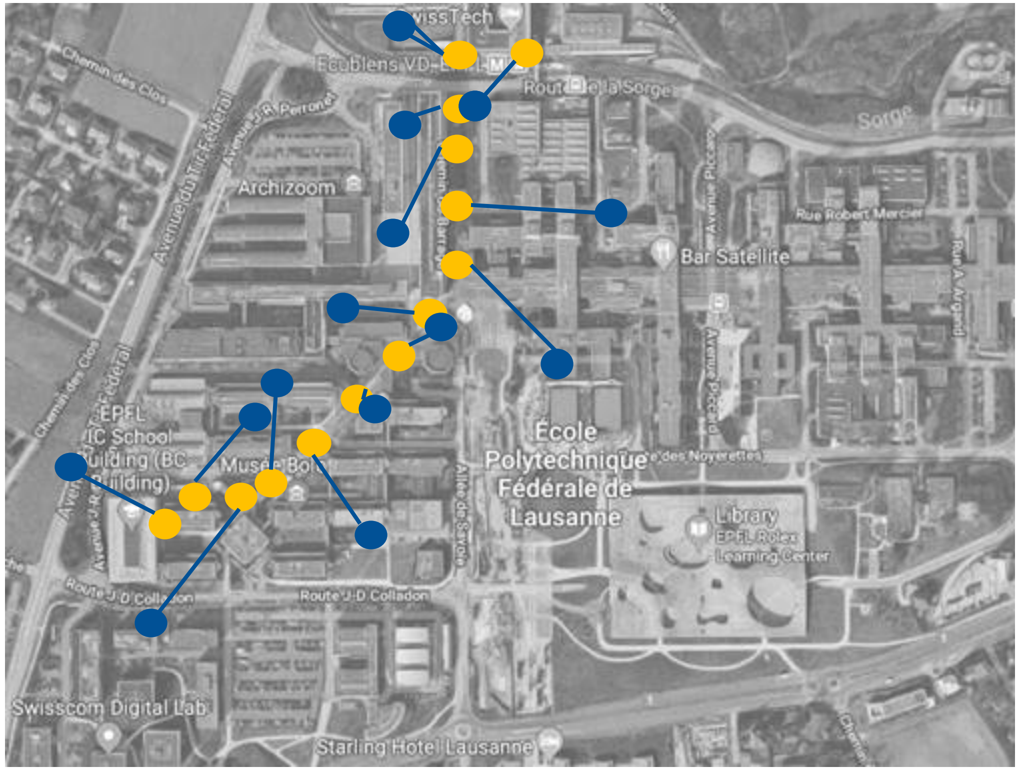
Significant privacy vs. utility trade-off
[Oya et al 2017b]

ϵ ↑

ϵ ↓



Add 2-dimensional
ϵ-differential privacy noise



How to protect location privacy

Perturbation

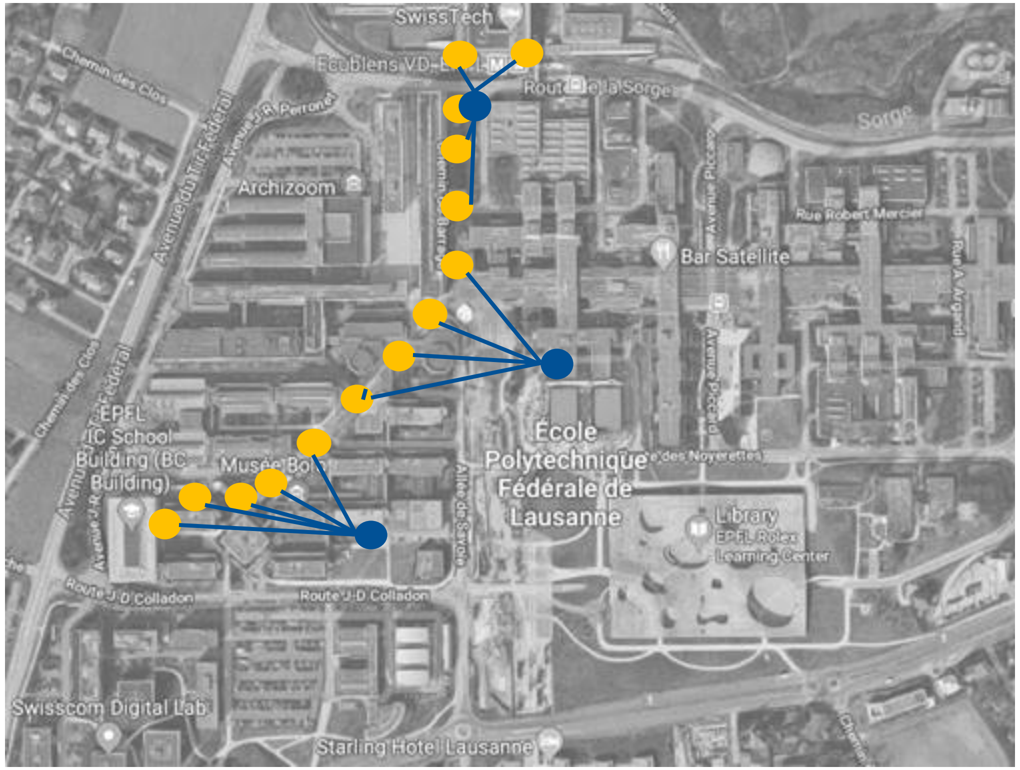
Spatial Obfuscation: Perturbation of locations using noise [Duckham & Kulik 2005]

Geo-indistinguishability [Andres et al. 2013]

As with differential privacy, we have
sequential composition: protection decreases
linearly with every sample → privacy degrades quickly

Release Geo-indistinguishability [Chatzikokolakis et al. 2014]

only draw noise when needed to keep utility
(i.e., when moving far from previous sample)



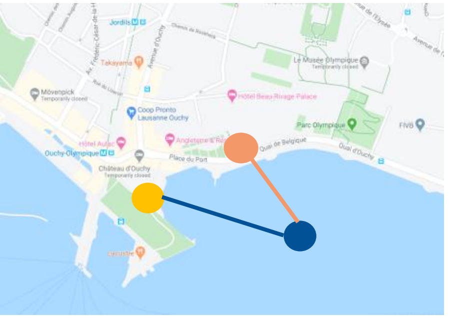
How to protect location privacy
Perturbation

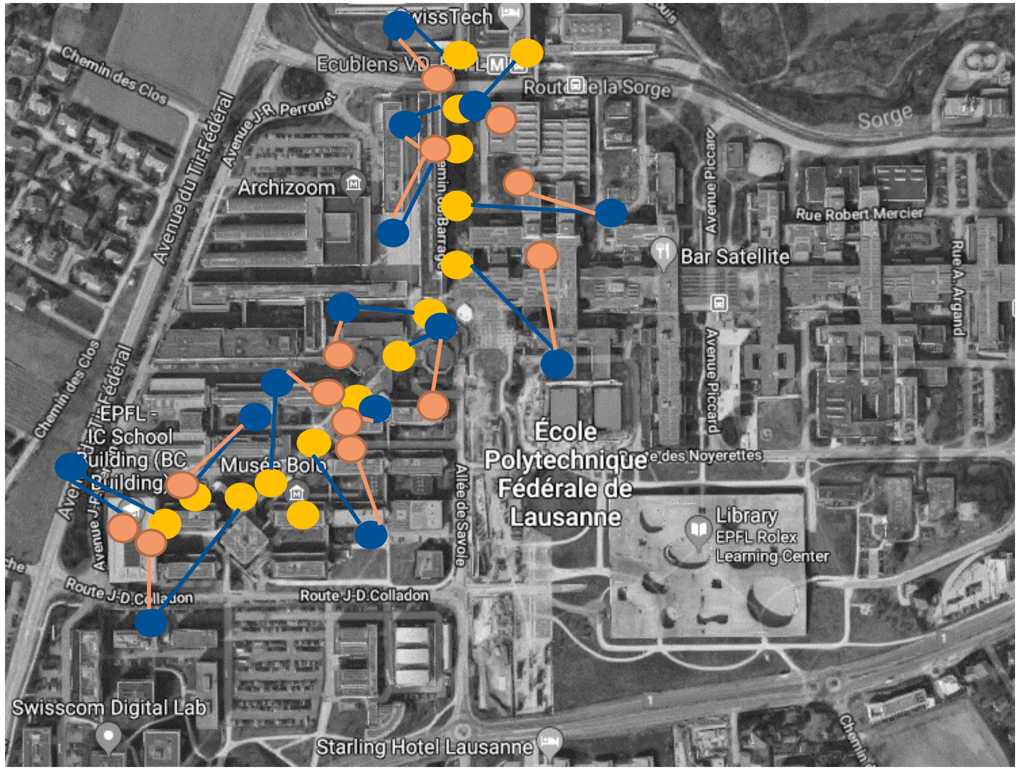
Spatial Obfuscation: Perturbation of locations using noise [Duckham & Kulik 2005]

Geo-indistinguishability [Andres et al. 2013]

Optimal remapping [Chatzikokolakis et al 2017] [Oya et al 2017]

Choose the best of the geo-indistinguishable options

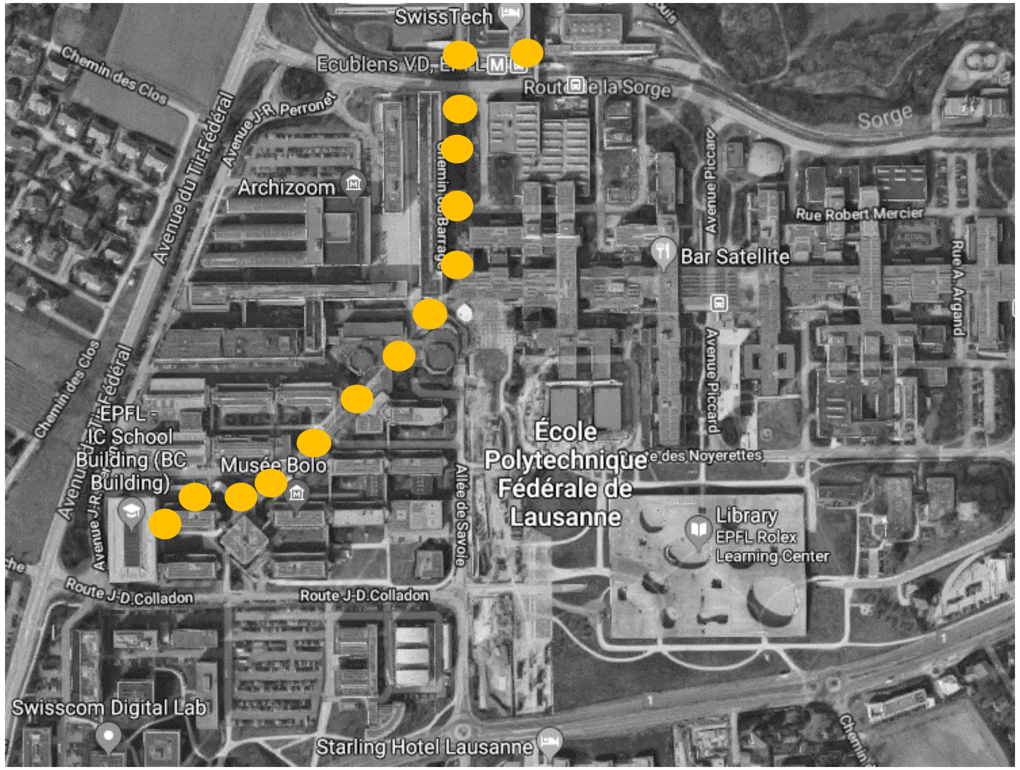




Requires a prior distribution to decide what's "best"!

How to protect location privacy
Hiding

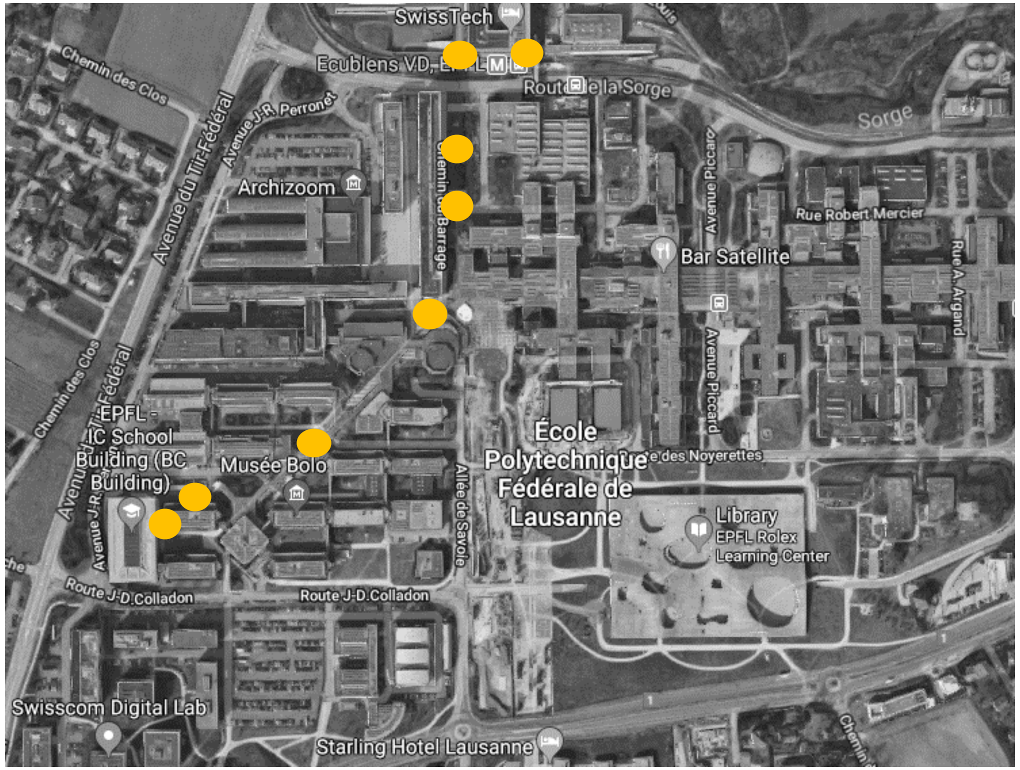
Hiding: Do not report some locations [Huang 2006][Hoh 2007]



How to protect location privacy
Hiding

Hiding: Do not report some locations [Huang 2006][Hoh 2007]

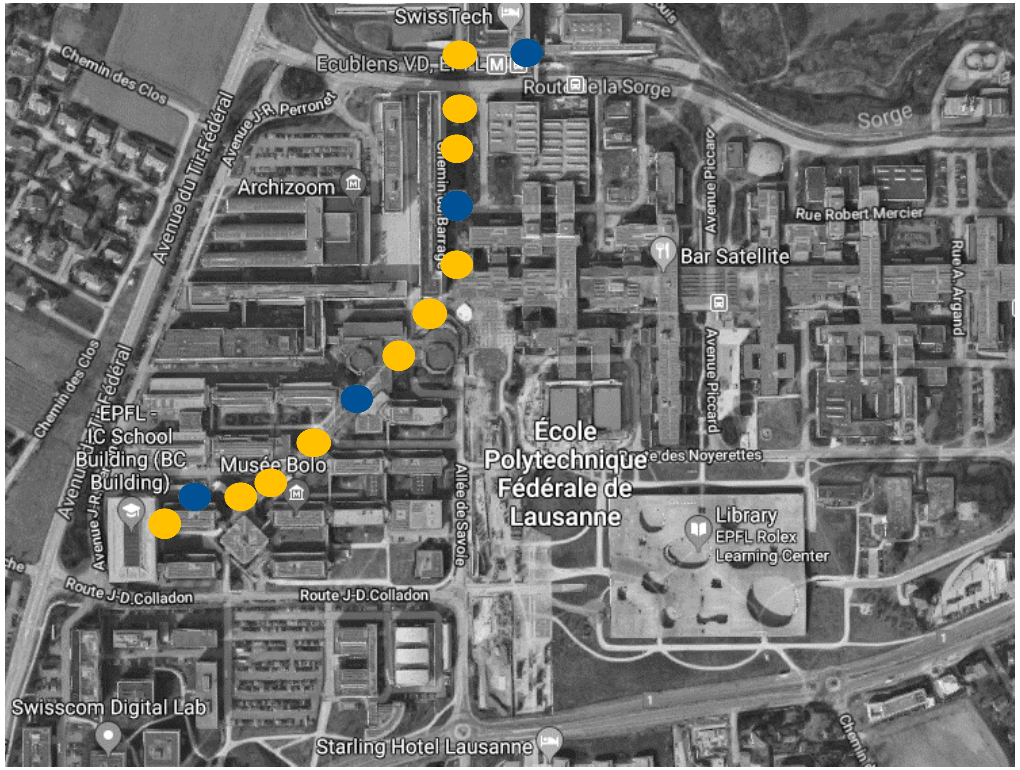
Random Hiding: Reveal a percentage of the points chosen at random (e.g, 50%)



How to protect location privacy
Hiding

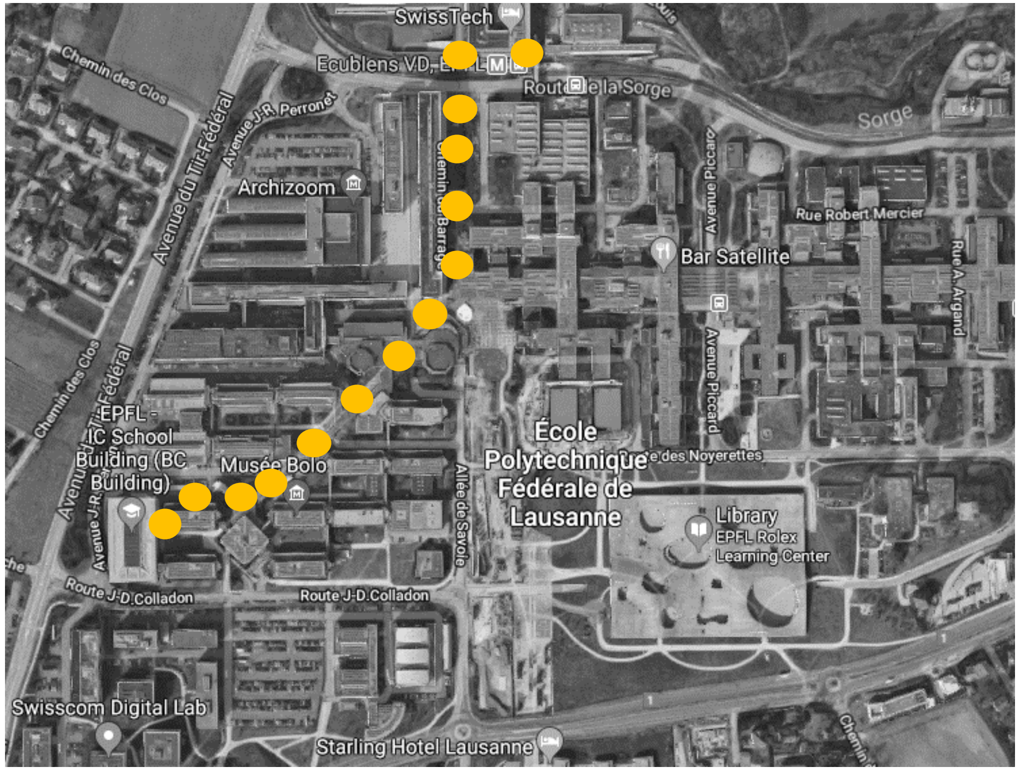
Hiding: Do not report some locations [Huang 2006][Hoh 2007]

Release: Reveal points only when needed



How to protect location privacy
Generalisation

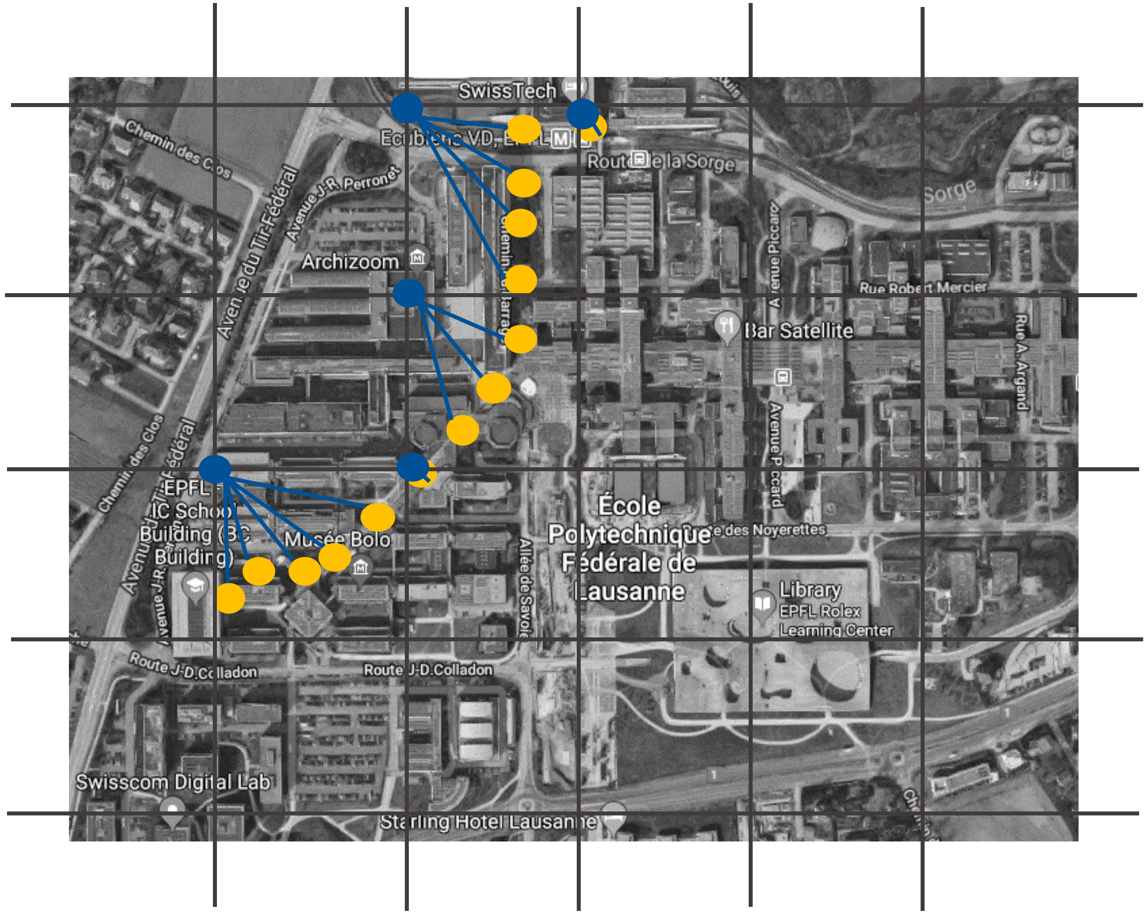
Generalization: reduce the precision of the reported locations [Bamba et al 2008]



How to protect location privacy
Generalisation

Generalization: reduce the precision of the reported locations [Bamba et al 2008]

Discretization: Map to grid points (Rounding - Floor) [Krumm 2009]



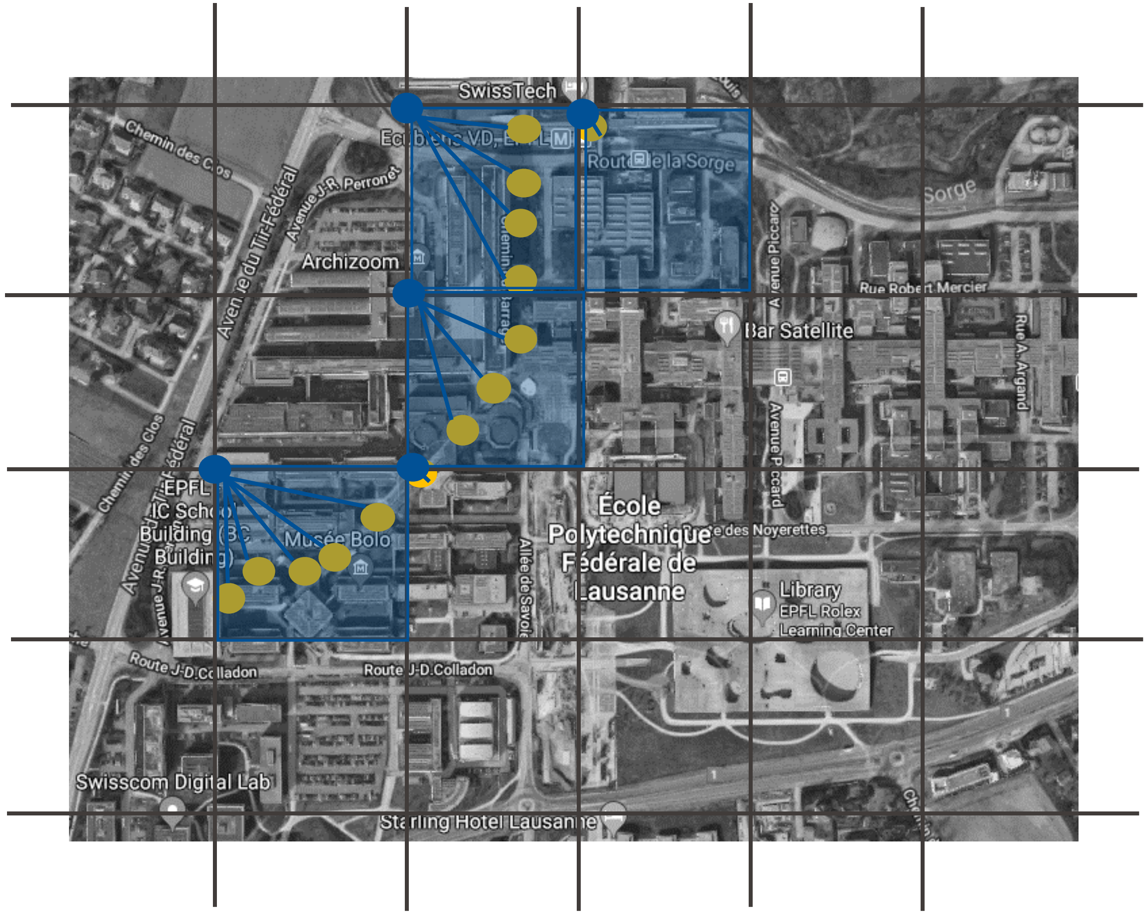
How to protect location privacy
Generalisation

Generalization: reduce the precision of the reported locations [Bamba et al 2008]

Discretization: Map to grid points (Rounding - Floor) [Krumm 2009]

Cloaking: Reveal a region

Fixed cloaks: always map to the same cloak



How to protect location privacy
Generalisation

Generalization: reduce the precision of the reported locations [Bamba et al 2008]

Discretization: Map to grid points (Rounding - Floor) [Krumm 2009]

Cloaking: Reveal a region

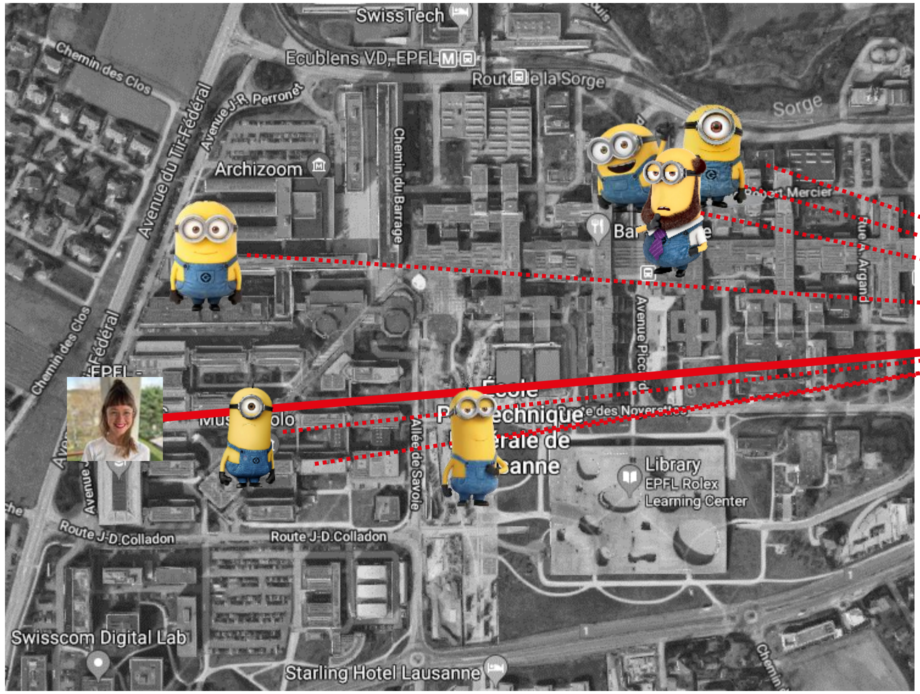
Fixed cloaks: always map to the same cloak

Location-dependent cloaks (centered on location)

k-anonymity based

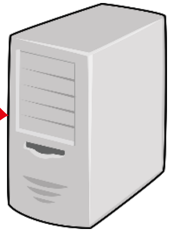
How to protect location privacy
A cautionary note on k-anonymity cloaking

[Gruteser & Grunwald 2003] and a long, long, long list of follow-up works



(x,y,Q) where (x,y) is the location and Q a query

(x,y,Q)



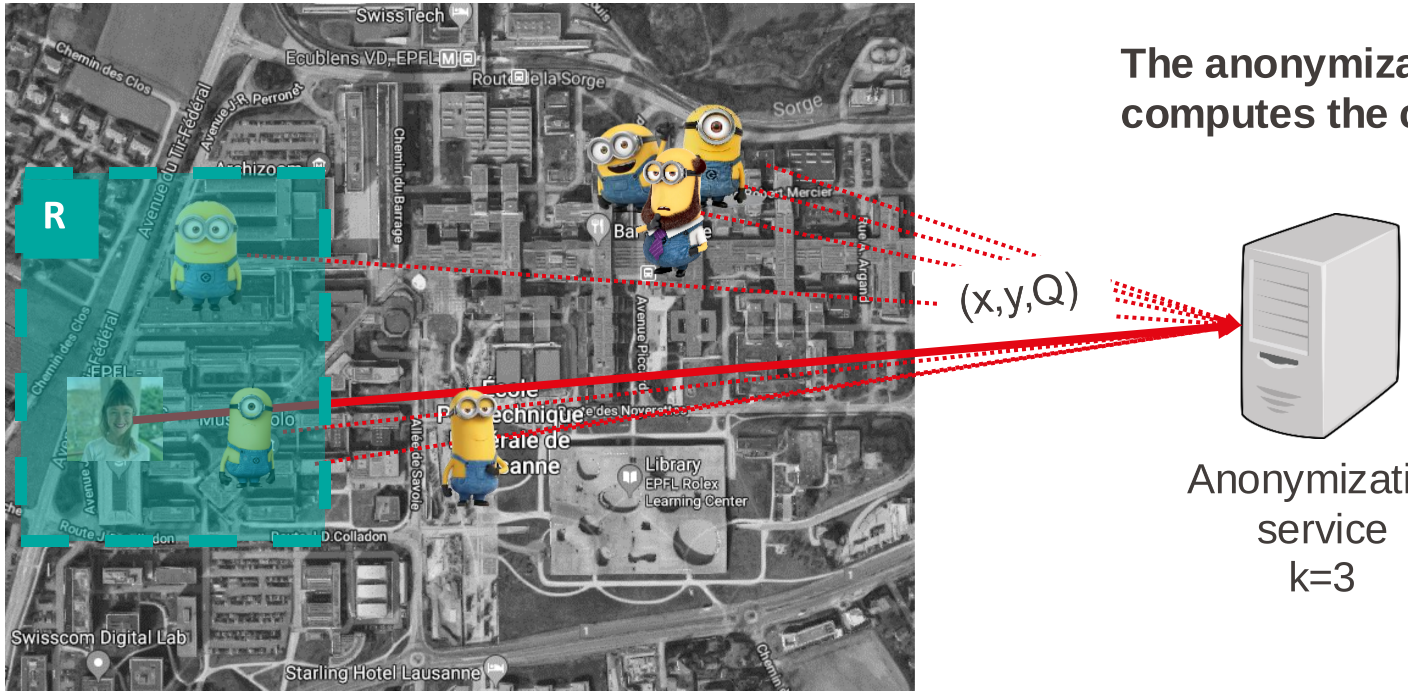
Anonymization
service
k=3

Privacy parameter

How to protect location privacy
A cautionary note on k-anonymity cloaking

[Gruteser & Grunwald 2003] and a long, long, long list of follow-up works

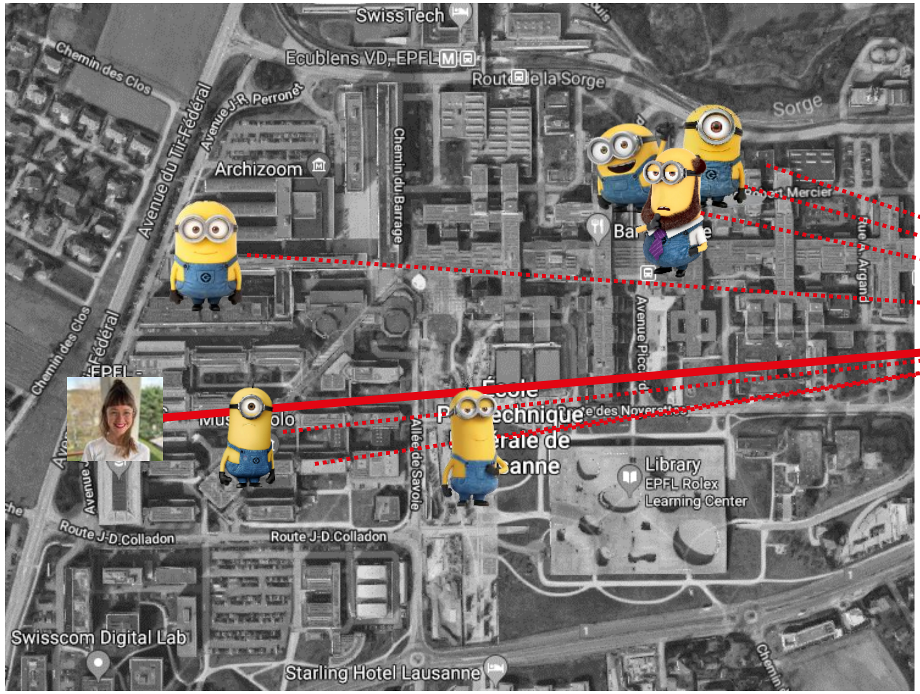
The anonymization service computes the cloak R



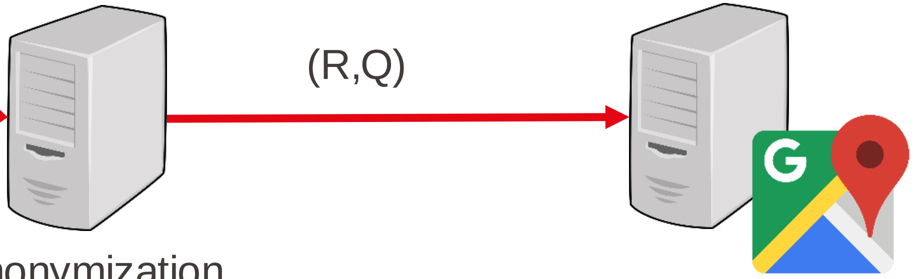
Anonymization
service
k=3

How to protect location privacy
A cautionary note on k-anonymity cloaking

[Gruteser & Grunwald 2003] and a long, long, long list of follow-up works



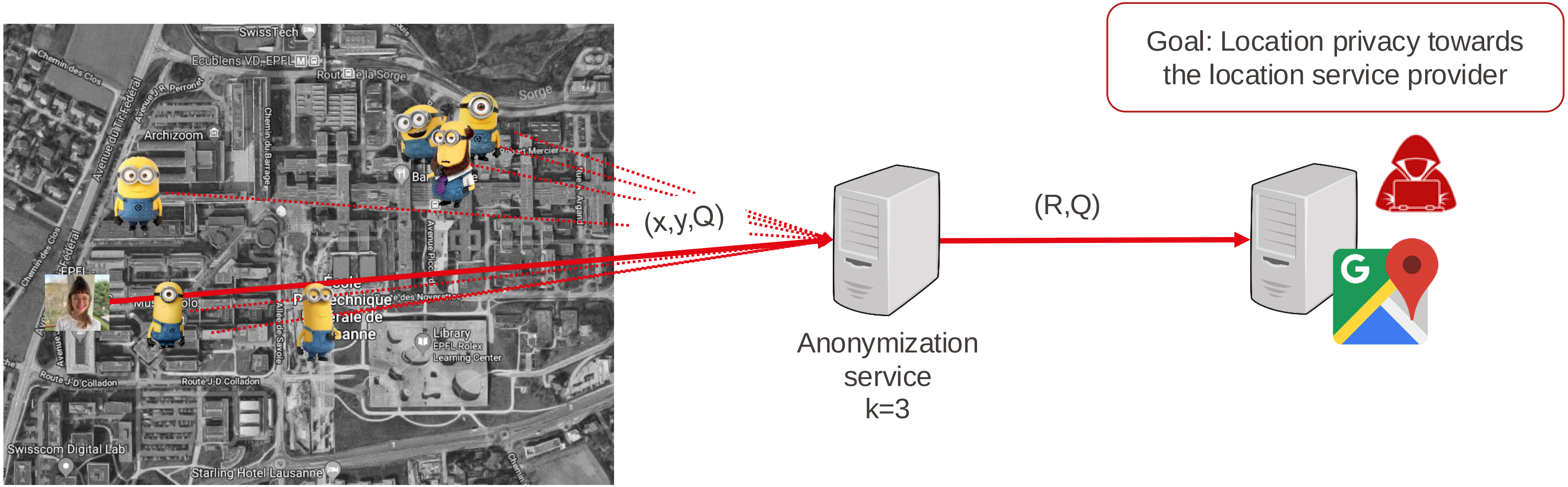
And sends the cloak and query to the location service



Anonymization service
k=3

How to protect location privacy
A cautionary note on k-anonymity cloaking

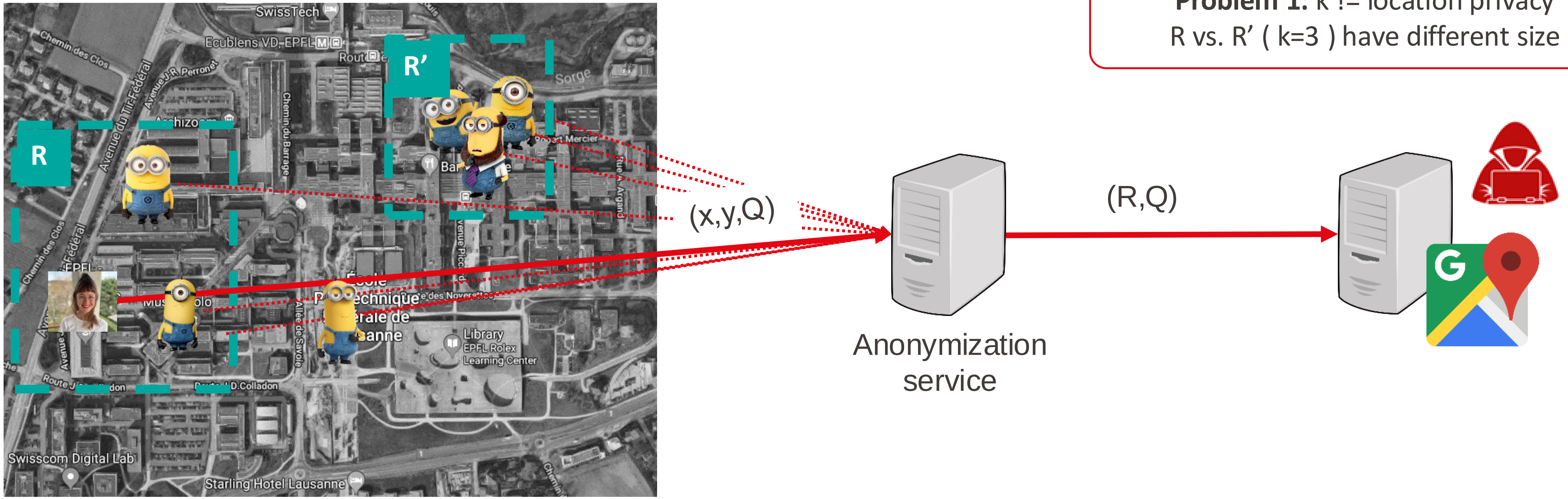
[Gruteser & Grunwald 2003] and a long, long, long list of follow-up works



How to protect location privacy
A cautionary note on k-anonymity cloaking

[Gruteser & Grunwald 2003] and a long, long, long list of follow-up works

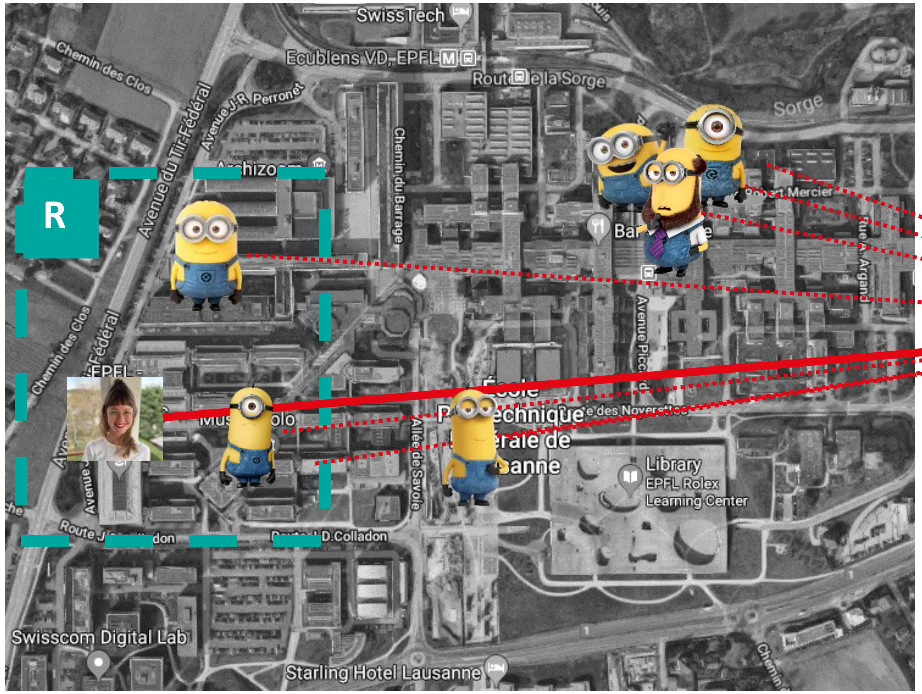
Problem 1: k != location privacy
R vs. R' (k=3) have different size

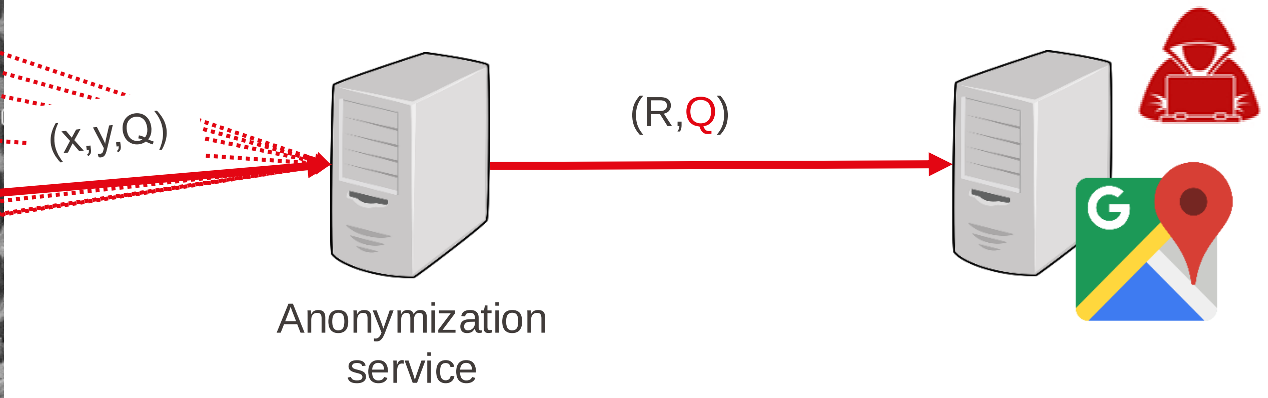


How to protect location privacy
A cautionary note on k-anonymity cloaking

[Gruteser & Grunwald 2003] and a long, long, long list of follow-up works

Problem 2: If service provider knows location, e.g., from query metadata, we have anonymity but not location privacy!!

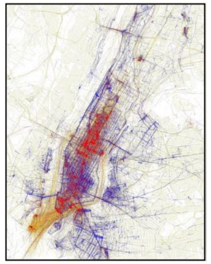


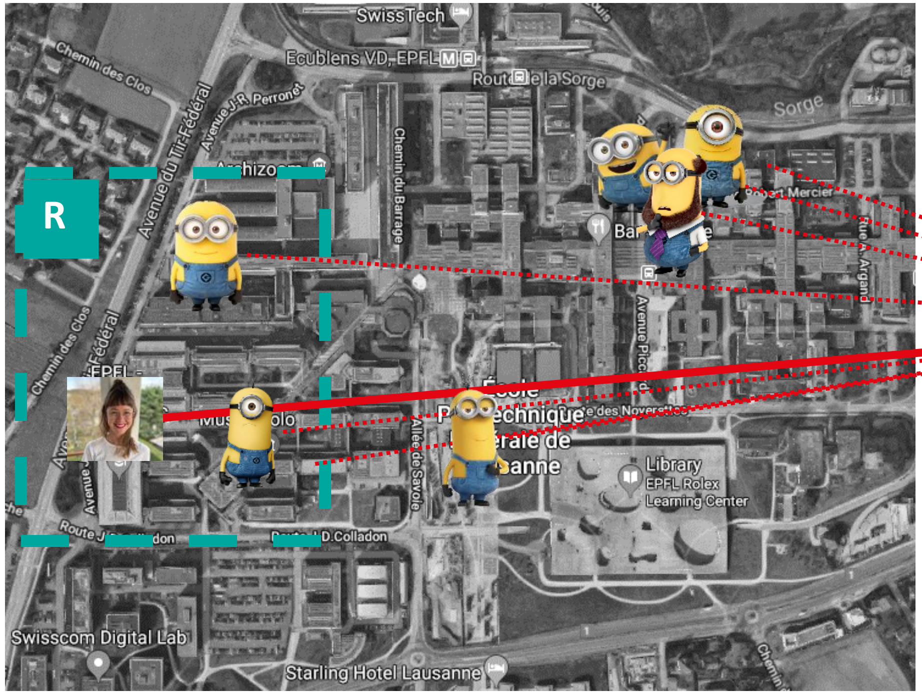


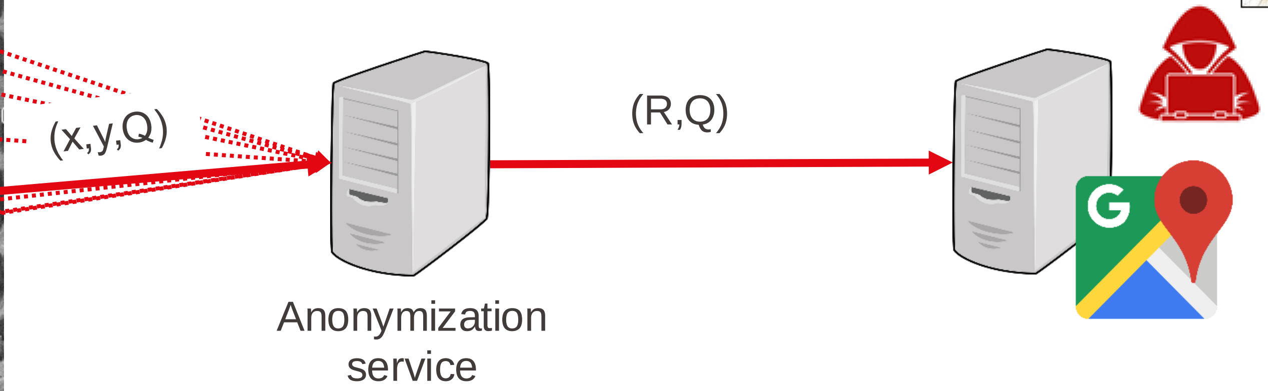
How to protect location privacy
A cautionary note on k-anonymity cloaking

[Gruteser & Grunwald 2003] and a long, long, long list of follow-up works

Problem 3: If service provider knows statistical information, e.g., public data, location privacy does not depend on people’s actual location!!

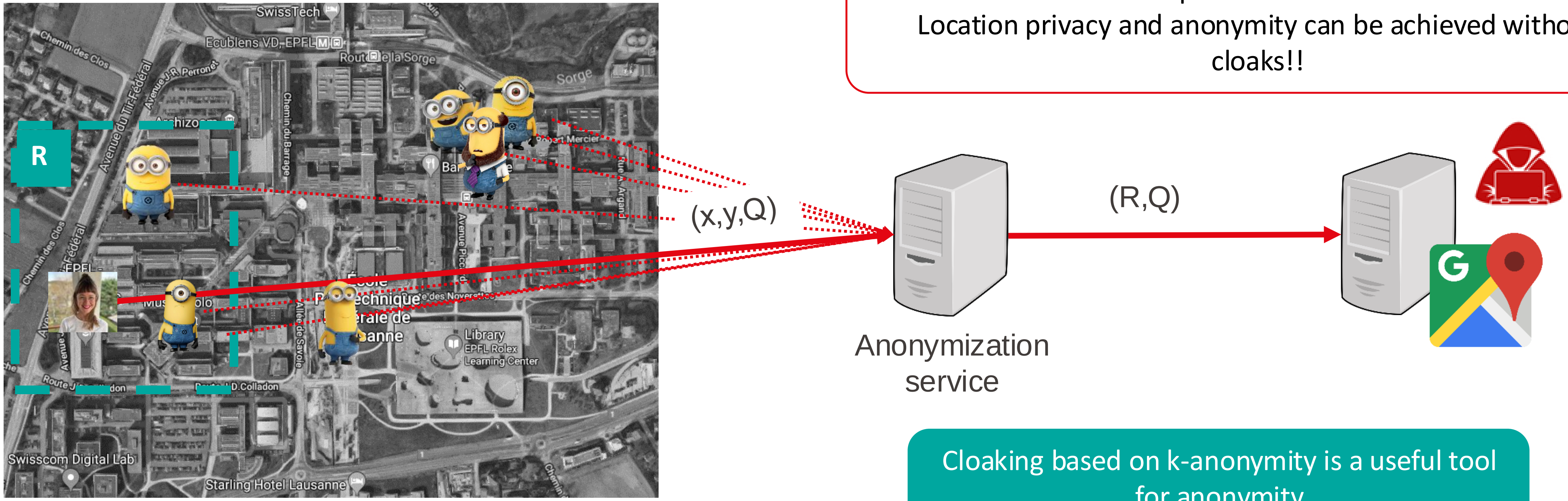






How to protect location privacy
A cautionary note on k-anonymity cloaking

[Gruteser & Grunwald 2003] and a long, long, long list of follow-up works



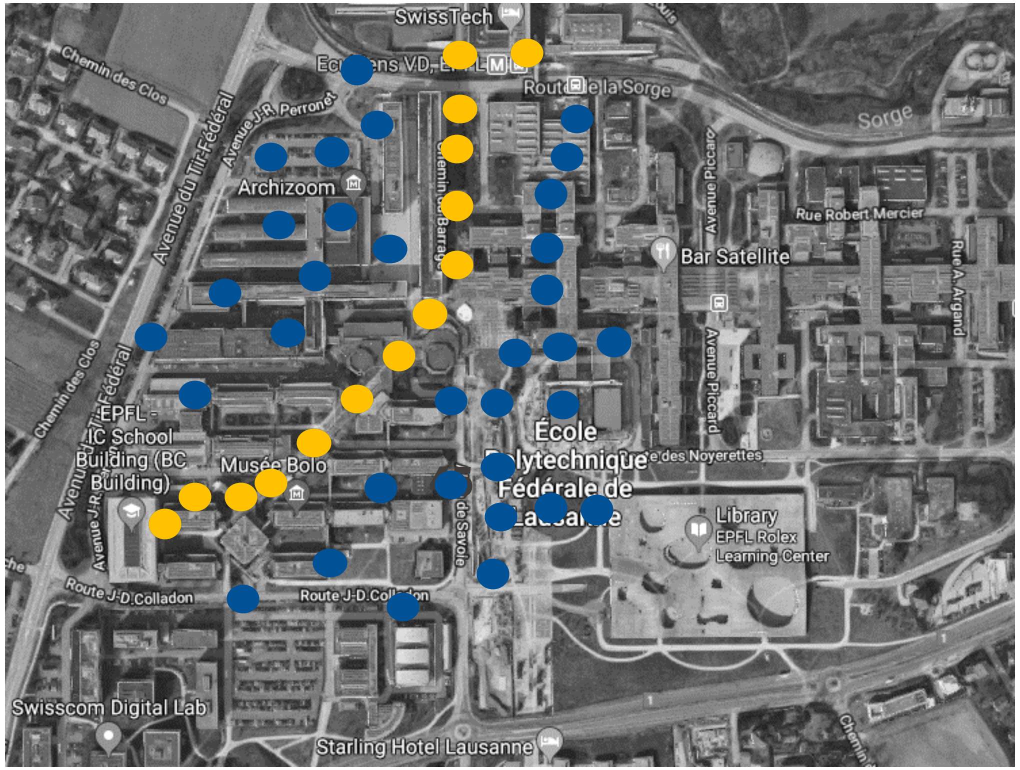
Problem 4: If the service provider has no additional knowledge, Location privacy and anonymity can be achieved without cloaks!!

Cloaking based on k-anonymity is a useful tool for anonymity
not location privacy

How to protect location privacy
Adding dummies

Dummy Locations: add decoy locations [Meyerovitz & Choudhury 2009]

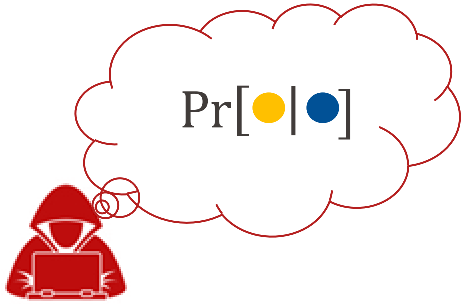
Difficult to create plausible dummies
[Chow & Golle 2009]



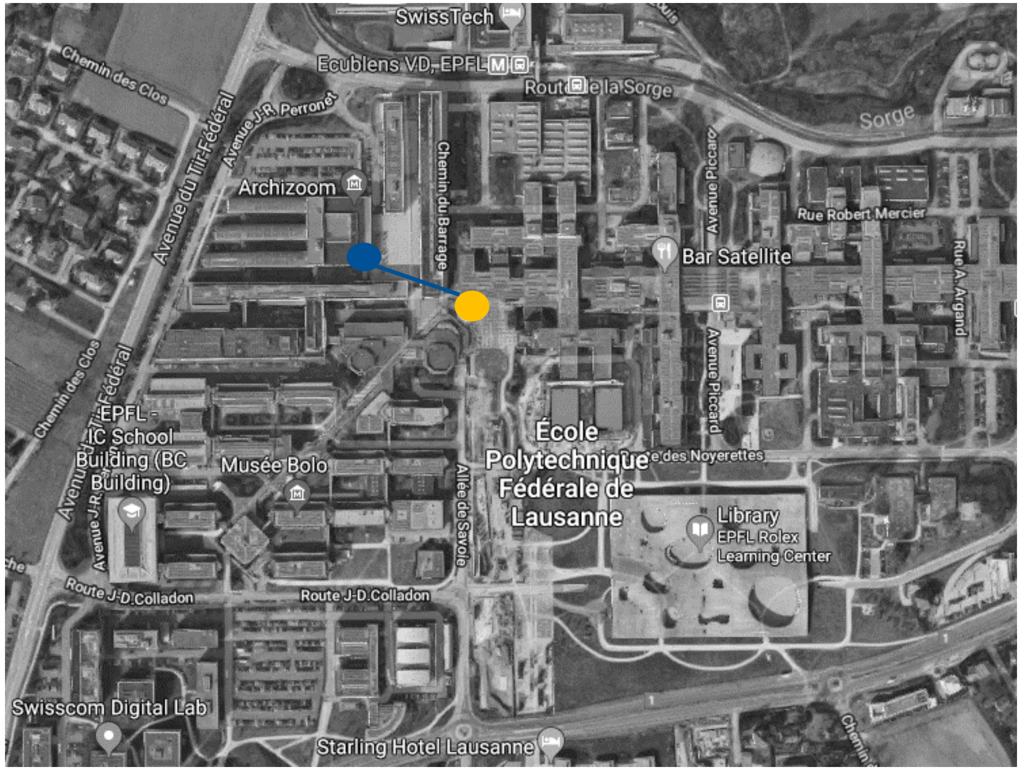
How to measure location privacy

How can we measure location privacy?
Strategic adversary

Strategic adversary
- Knows the defense mechanism
  $\bullet = K(\bullet)$
- Given released location, estimates most likely real location



Computing this probability is hard for location traces: too many plausible options.
→ Use sampling methods (MCMC)



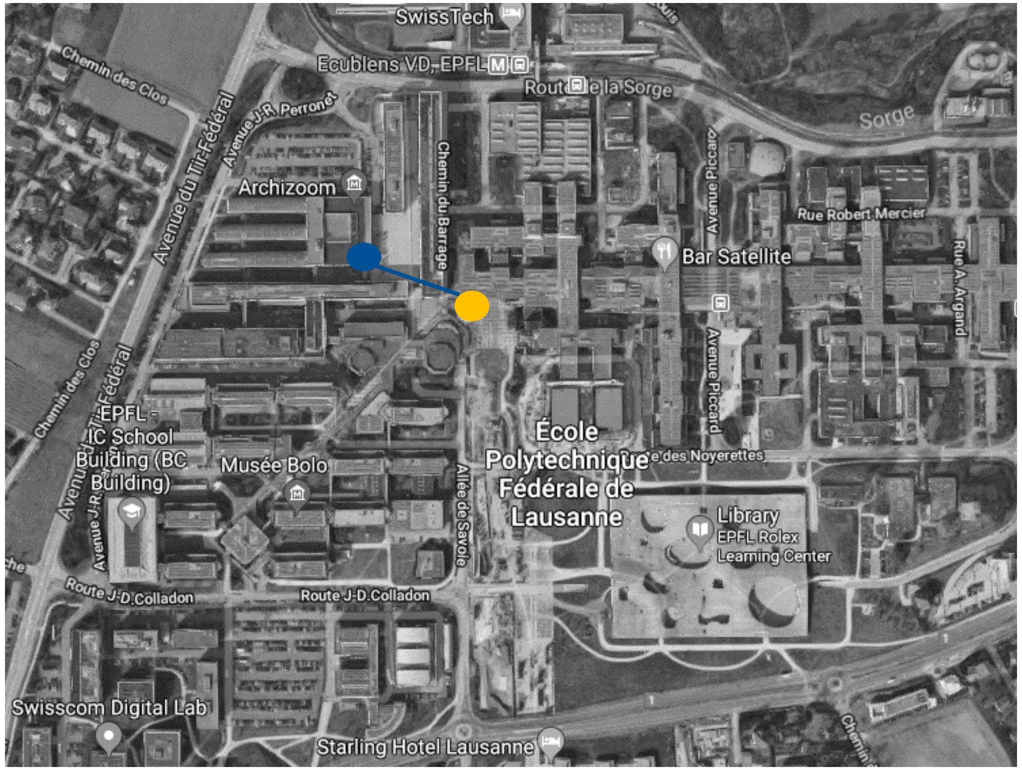
How can we measure location privacy?
Privacy error

Privacy error

Accuracy: how much variance in estimation
- Confidence interval

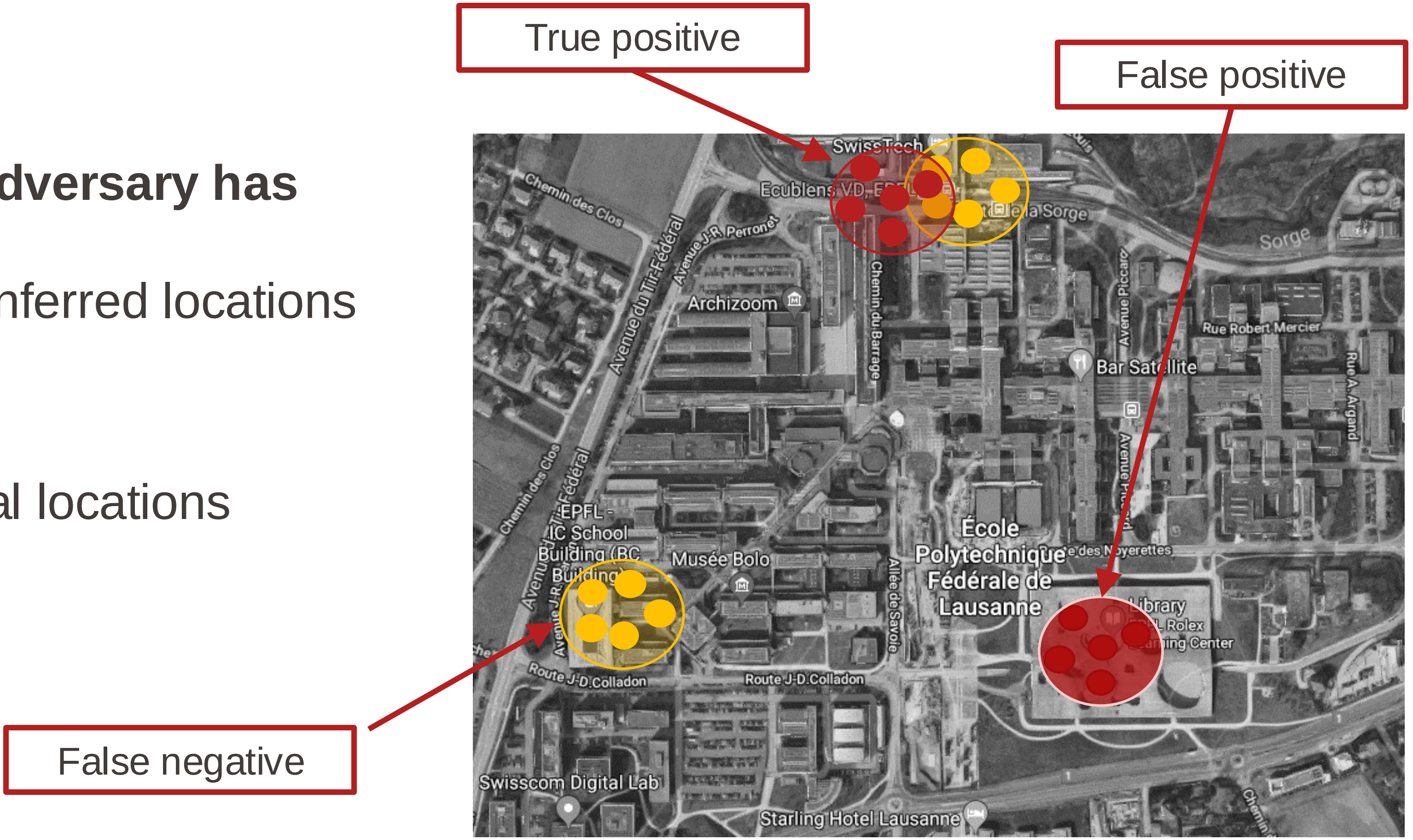
Correctness: how close to reality
- Adversary's error [Shokri et al 2011]

Certainty: how sure of the guess
- Entropy [Oya et al 2017]



How can we measure location privacy?

Privacy error



Privacy is achieved if the adversary has

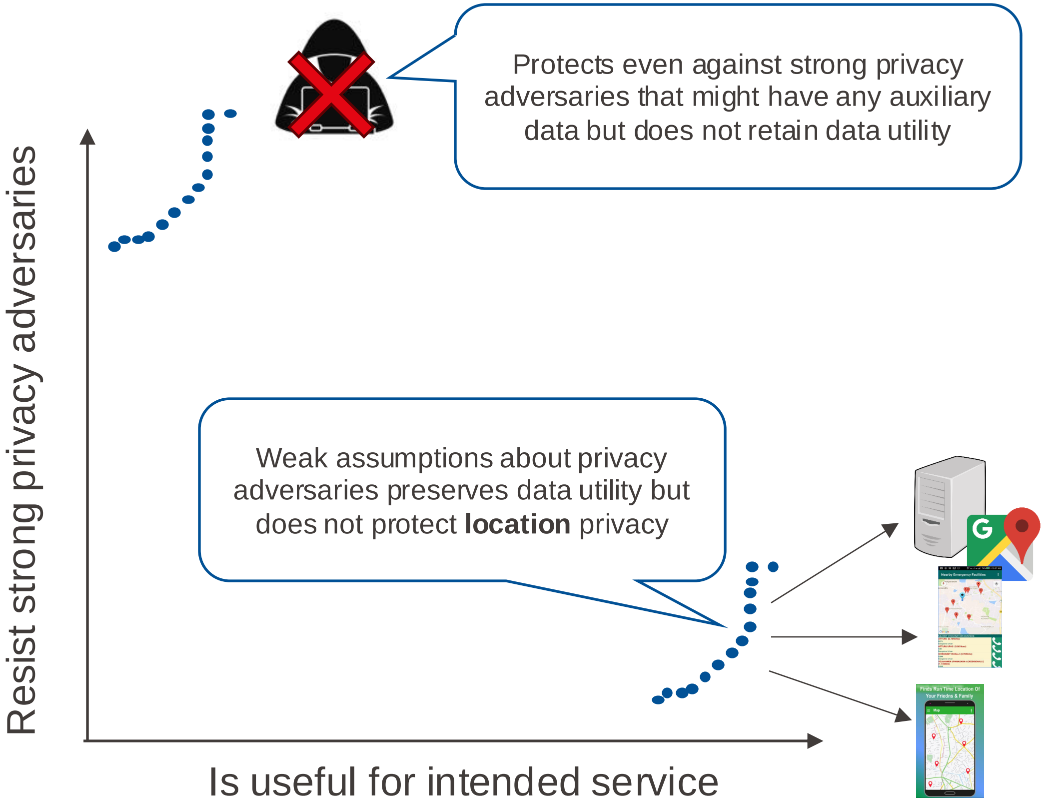
Low precision: many false inferred locations

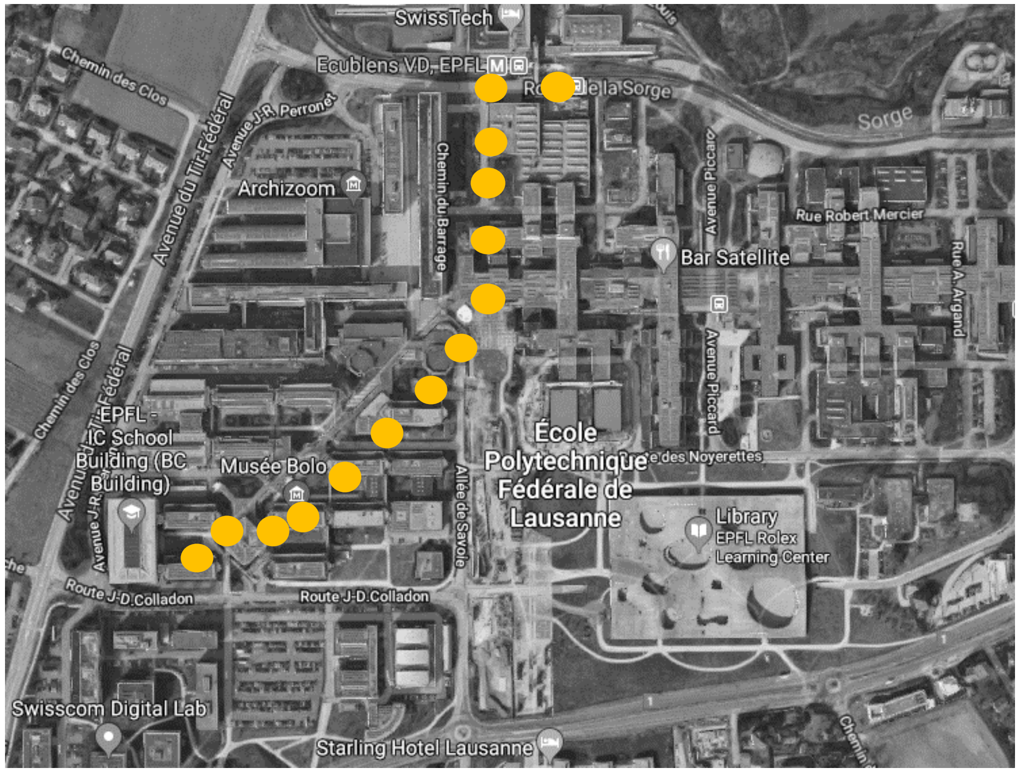
$$Precision = \frac{TP}{TP+FP}$$

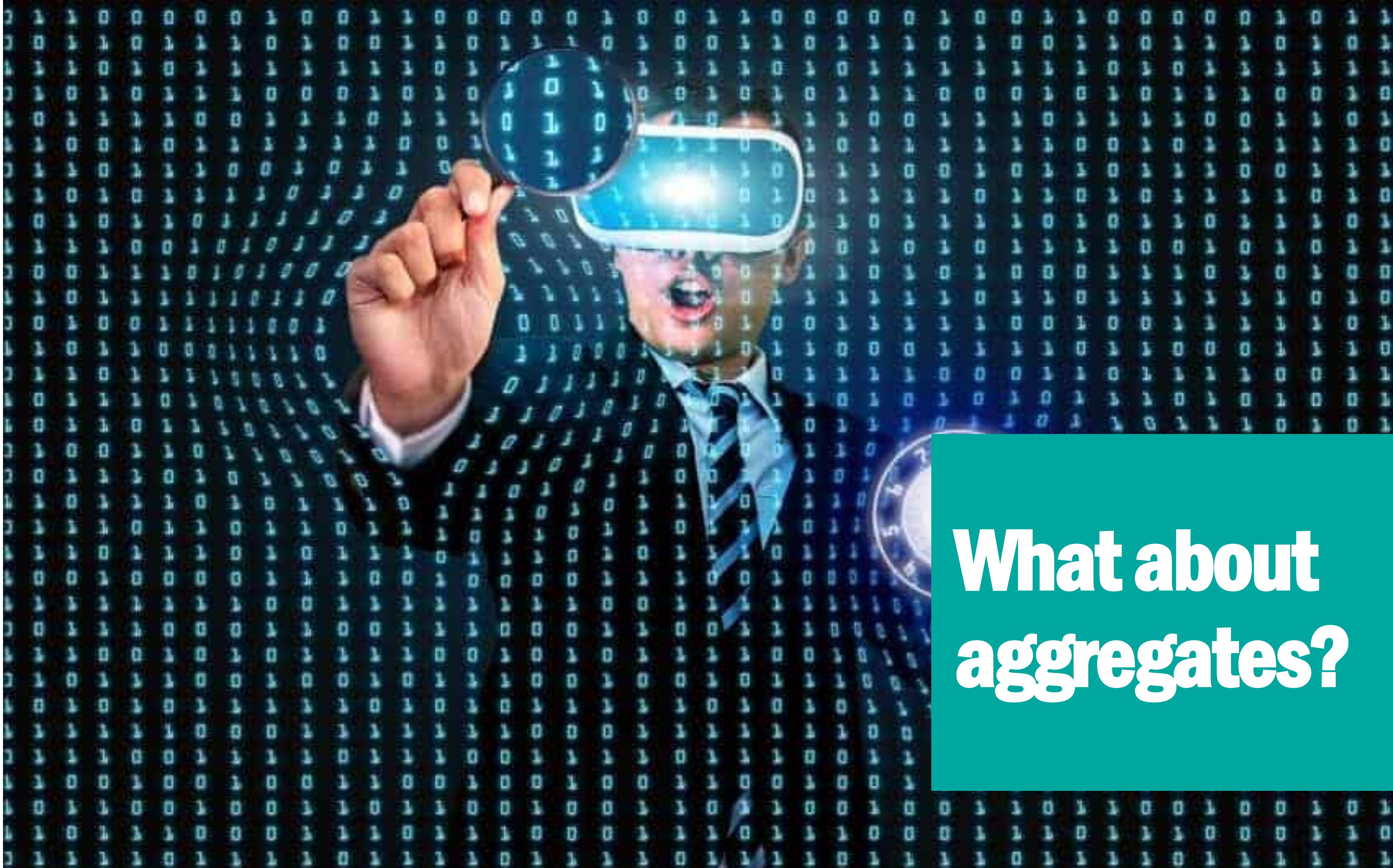
Low recall: misses many real locations

$$Recall = \frac{TP}{TP+FN}$$

If we use these measures to assess the protections…

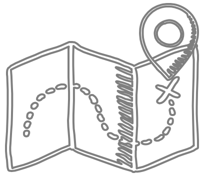


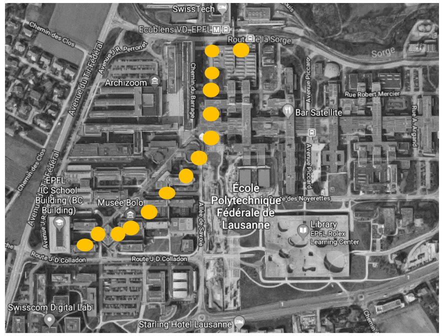




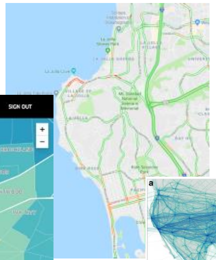
What about aggregates?

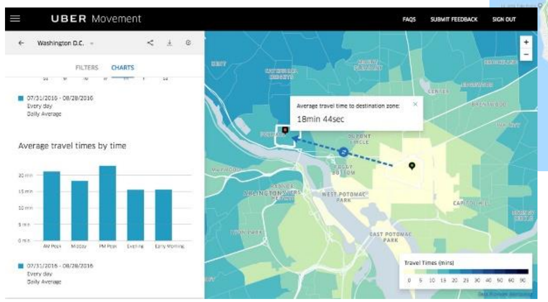
What about hiding in the crowd?
Aggregate statistics

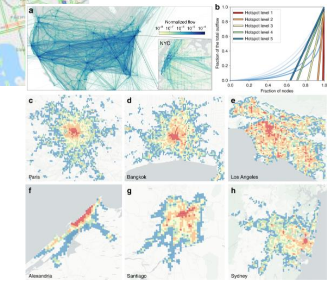




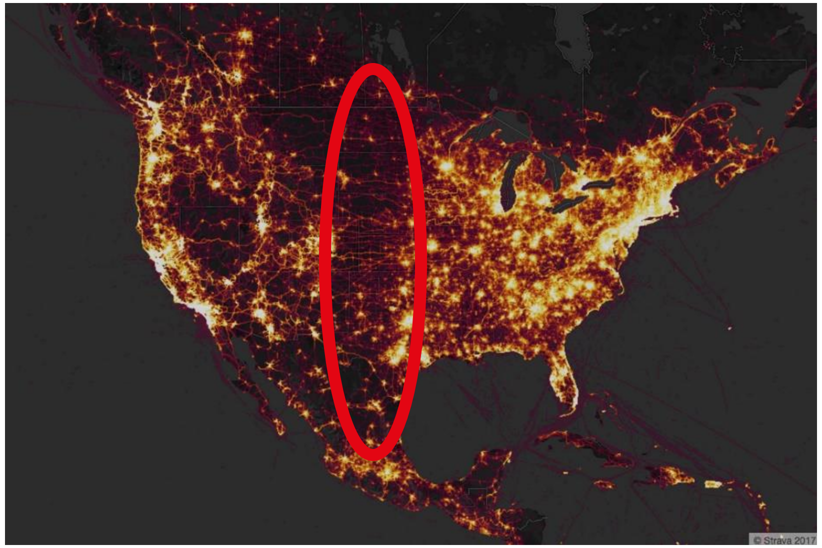
So would aggregates be secure?







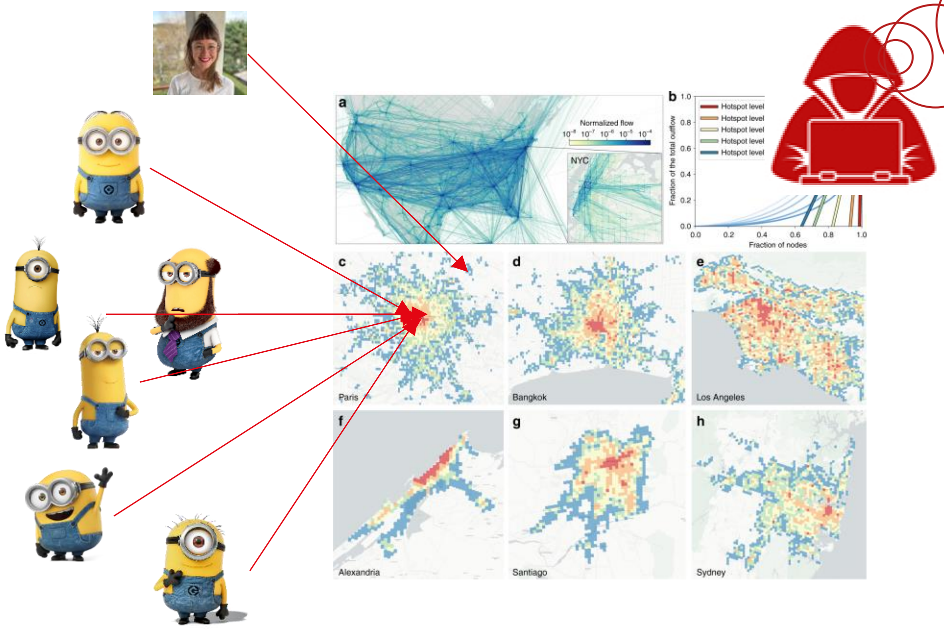
Even aggregate location is sensitive…



- https://www.wired.com/story/strava-heat-map-military-bases-fitness-trackers-privacy/

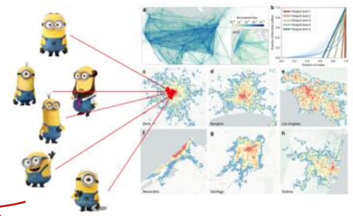
What can be inferred from aggregate location data?

Outliers create “particular” statistics

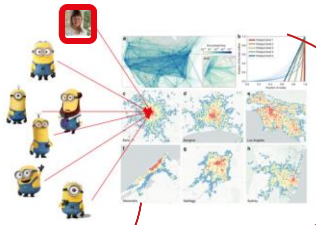


Aggregates reflect statistics of the population

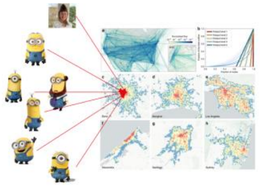
Train machine learning models to distinguish statistical patterns with/without outlier



vs.

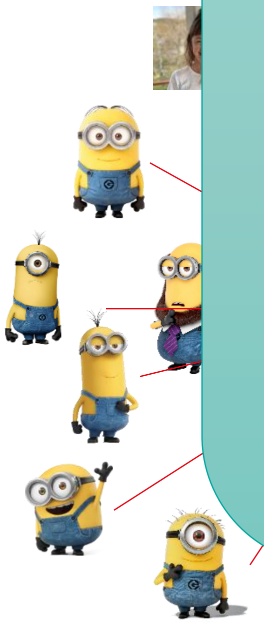


Once membership is known, aggregates enable further inferences



What can be inferred from aggregate location data?

Outliers cre



And this is very hard to defend against while keeping utility


nature COMMUNICATIONS

MATTERS ARISING

Check for updates

https://doi.org/10.1038/s41467-021-27566-0 OPEN

On the difficulty of achieving Differential Privacy in practice: user-level guarantees in aggregate location data

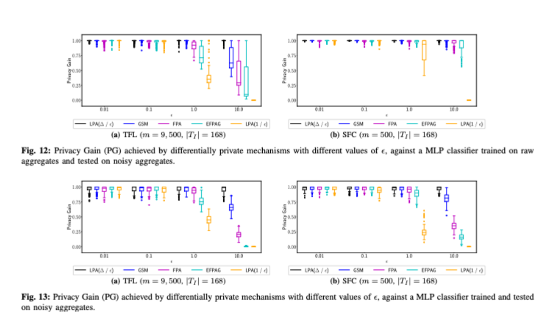
Florimond Houssiau[1,2], Luc Rocher[1,3] & Yves-Alexandre de Montjoye[1]


Knock Knock, Who's There?
Membership Inference on Aggregate Location Data*

Apostolos Pyrgelis, University College London, apostolos.pyrgelis.14@ucl.ac.uk

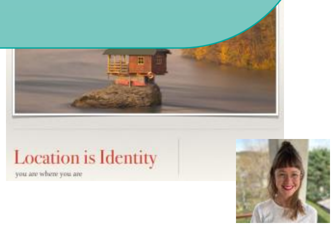
Carmela Troncoso, IMDEA Software Institute, carmela.troncoso@imdea.org

Emiliano De Cristofaro, University College London, e.decristofaro@ucl.ac.uk



ish

er inferences



Aggregates reflect statistics of the population

Take away

Take aways

- Location data contains a lot of sensitive information about us
  - About our health status, our religious beliefs, our financial situation, whom we interact with
- Simple inference attacks can extract this information
- Hard to protect location data against inference attacks while preserving its utility
  - Techniques like perturbation, generalisation, dummies, hiding all come with stringent privacy utility trade-offs
  - Aggregation is a weak privacy-preserving mechanism: **membership attacks** are feasible

→ To design effective defenses, we need to adjust them to the adversarial model

References

[Andres et al 2013] M. E. Andres, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in CCS, 2013.

[Ashbrook & Starner 2003] D. Ashbrook and T. Starner. Using GPS to learn significant locations and predict movement across multiple users. Personal and Ubiquitous Computing, 7(5):275–286, 2003.

[Backes et al 2017] M. Backes, M. Humbert, J. Pang, and Y. Zhang, "walk2friends: Inferring social links from mobility profiles," in CCS, 2017.

[Bamba et al 2008] B. Bamba, L. Liu, P. Pesti, and T. Wang, "Supporting anonymous location queries in mobile environments with PrivacyGrid," in WWW, 2008.

[Bilogrevic et al 2015] Bilogrevic, I., Huguenin, K., Mihaila, S., Shokri, R., & Hubaux, J. P. Predicting users' motivations behind location check-ins and utility implications of privacy protection mechanisms. In NDSS 2015.

[Bindschaedler & Shokri 2016] V. Bindschaedler and R. Shokri, "Synthesizing plausible privacy-preserving location traces," in IEEE S&P, 2016.

[Chatzikokolakis et al 2014] K. Chatzikokolakis, C. Palamidessi, and M. Stronati, "A predictive differentially-private mechanism for mobility traces," in PETS, 2014.

[Chatzikokolakis et al 2017] K. Chatzikokolakis, E. Elsalamouny, and C. Palamidessi, "Efficient utility improvement for location privacy," PETS, 2017.

[Cho et al 2100] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: User movement in location-based social networks. In KDD, 2011.

[Chow & Golle 2009] Chow, R., & Golle, P. Faking contextual data for fun, profit, and privacy. In WPES 2009.

[Crandall et al 2010] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg, "Inferring social ties from geographic coincidences," PNATL ACAD SCI USA, 2010.

References

[De Montjoye et al 2013] De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. Unique in the crowd: The privacy bounds of human mobility. Scientific reports, 3, 1376. 2013.

[De Montjoye et al 2015] De Montjoye, Y. A., Radaelli, L., & Singh, V. K. Unique in the shopping mall: On the reidentifiability of credit card metadata. Science, 2015.

[Duckham & Kulik 2005] M. Duckham and L. Kulik. A formal model of obfuscation and negotiation for location privacy. In Pervasive Computing, 2005.

[Dwork et al 2010] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum. Differential privacy under continual observation. In STOC, 2010

[Eagle et al 2009] N. Eagle, A. S. Pentland, and D. Lazer, “Inferring friendship network structure by using mobile phone data,” P NATL ACAD SCI USA, 2009

[Ester 1996] M. Ester, H.-P. Kriegel, J. Sander, X. Xu et al., “A density-based algorithm for discovering clusters in large spatial databases with noise,” in KDD, 1996.

[Felbo et al 2017] Felbo, B., Sundsøy, P., Lehmann, S., & de Montjoye, Y. A. Modeling the Temporal Nature of Human Behavior for Demographics Prediction. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2017.

[Gambs et al 2012] S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez, “Next place prediction using mobility markov chains,” in MPM, 2012.

[Golle & Partridge 2009] Golle, P., & Partridge, K. (2009, May). On the anonymity of home/work location pairs. In Pervasive, 2009

[Gruteser & Grunwald 2003] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In MobiSys, 2003.

[Hoh 2007] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, “Preserving privacy in GPS traces via uncertainty-aware path cloaking,” in CCS, 2007.

[Huang 2006] L. Huang, H. Yamane, K. Matsuura, and K. Sezaki, “Silent cascade: Enhancing location privacy without communication qos degradation,” in SPC. 2006.

[Krumm 2007] J. Krumm. Inference attacks on location tracks. In Pervasive, 2007.

References

[Krumm 2009] Krumm, J. (2009). A survey of computational location privacy. Personal and Ubiquitous Computing, 13(6), 391-399.

[Liao et al 2005] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition using relational Markov networks. In IJCAI, 2005.

[Liao et al 2007] L. Liao, D. J. Patterson, D. Fox, and H. Kautz. Learning and inferring transportation routines. Artificial Intelligence, (171):311–331, 2007

[Meyerovitz & Choudhury 2009] Meyerowitz, J., & Roy Choudhury, R. Hiding stars with fireworks: location privacy through camouflage. In Mobicomm 2009.

[Oya et al 2017] S. Oya, C. Troncoso, and F. Perez-Gonzalez, "Back to the drawing board: Revisiting the design of optimal location privacy-preserving mechanisms," in CCS, 2017.

[Oya et al 2017b] S. Oya, C. Troncoso, and F. Pérez-González. Is Geo-Indistinguishability What You Are Looking for? WPES 2017

[Pang & Zhang 2017] J. Pang and Y. Zhang, DeepCity: A Feature Learning Framework for Mining Location Check-Ins. ICWSM 2017

[Quercia el al 2011] D. Quercia, I. Leontiadis, L. McNamara, C. Mascolo, and J. Crowcroft. Spotme if you can: Randomized responses for location obfuscation on mobile phones. In ICDCS, 2011

[Pelleg & Moore 2000] D. Pelleg and A. Moore, "X-means: Extending k-means with efficient estimation of the number of clusters," ICML, 2000.

[Rastogi & Nath 2010] V. Rastogi and S. Nath. Differentially private aggregation of distributed time-series with transformation and encryption. In SIGMOD, 2010

[Shokri et al 2010] R. Shokri, C. Troncoso, C. Díaz, J. Freudiger, J.-P. Hubaux. Unraveling an old cloak. k-anonymity for location privacy. WPES 2010.

[Shokri et al 2011] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," IEEE S&P, 2011.

[Zhang & Bolot 2011] H. Zang and J. Bolot, "Anonymization of location data does not work: A large-scale measurement study," MobiCom, 2011.