AtlantTIC
Research Center for
Information & Communication Technologies

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

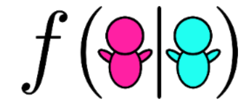
Rethinking Location Privacy for Unknown Mobility Behaviors

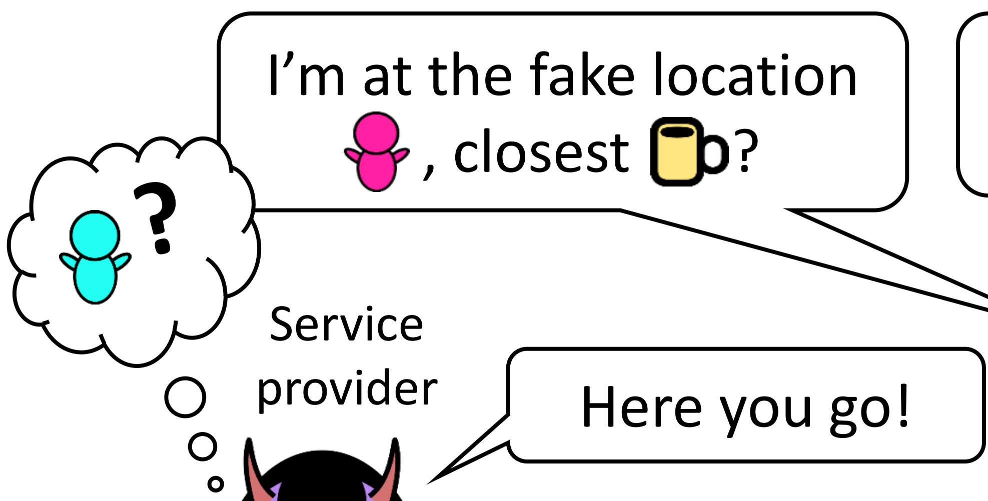
Simon Oya (UVigo)

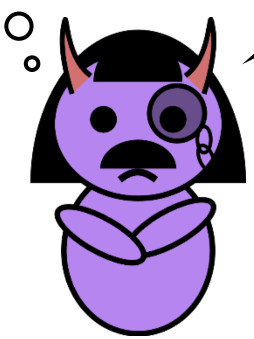
Carmela Troncoso (EPFL)

Fernando Pérez-González (UVigo)

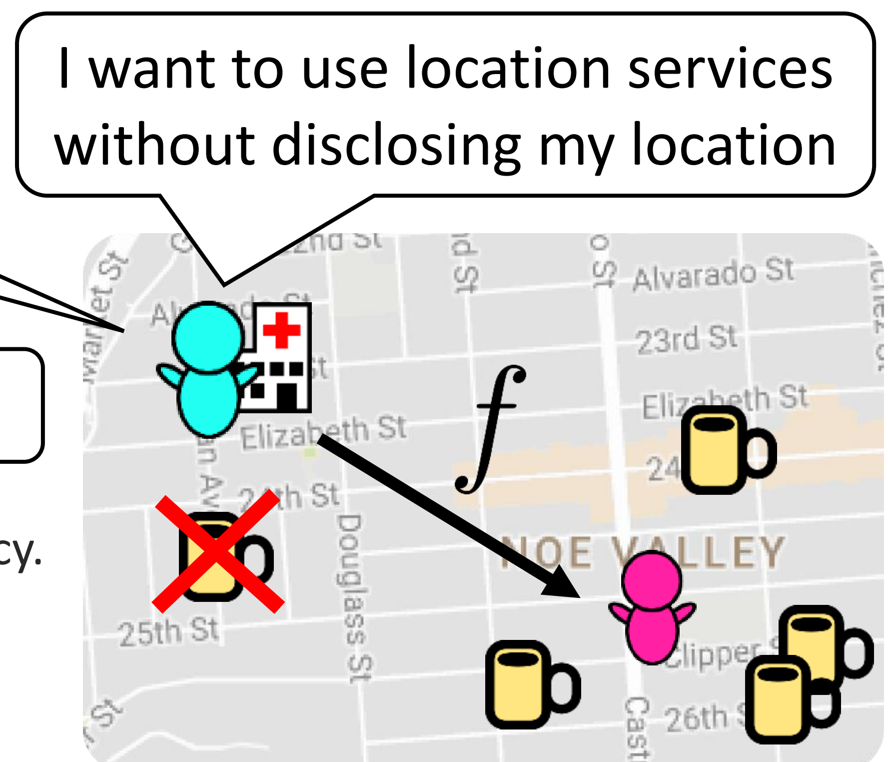
Motivation: Obfuscation-based Location Privacy

- Location information is sensitive.
- Location Privacy-Protection Mechanisms (LPPMs) 





- User gets some privacy.
- User loses some quality of service.



LPPM Design Notions: Metrics and Mobility Models

Quality Loss

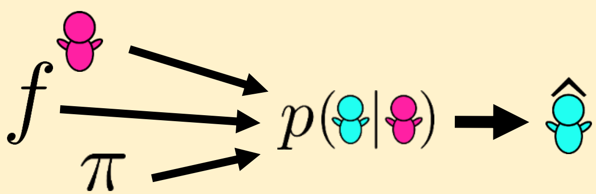
- Example: **Average Loss**

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

Euclidean, Hamming, semantic, …

Privacy

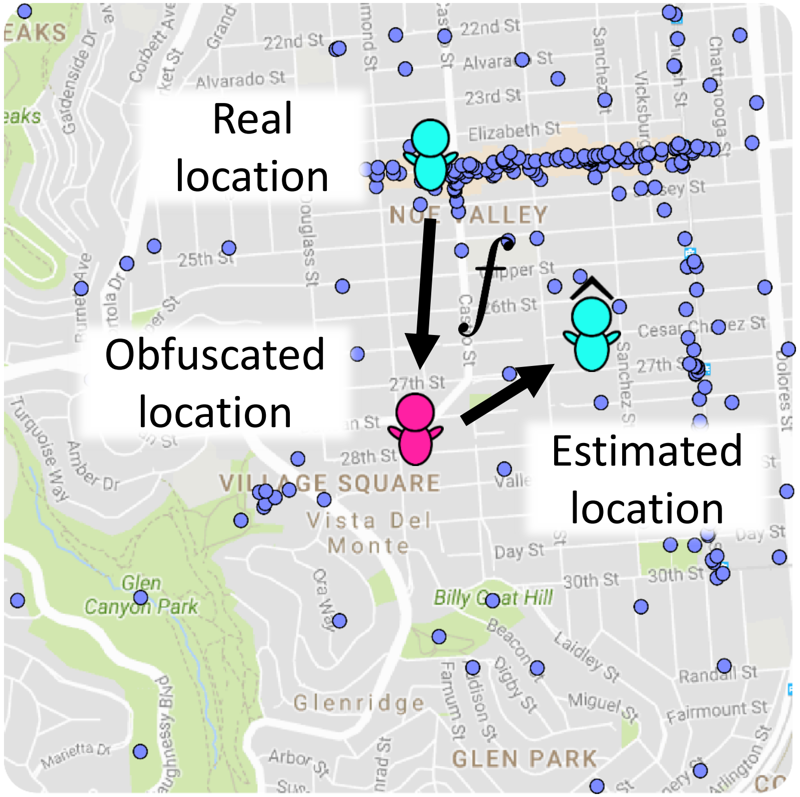
- Example: **Average Adversary Error**



Adversary's estimation of the real location

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

Euclidean, Hamming, semantic, …



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

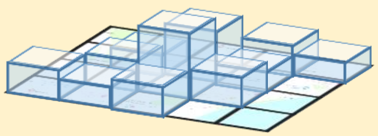
AtlantTIC Research Center for Information & Communication Technologies

Universida de Vigo

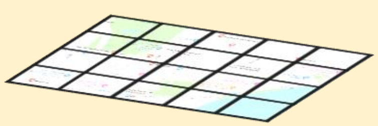
LPPM Design Notions: Metrics and Mobility Models

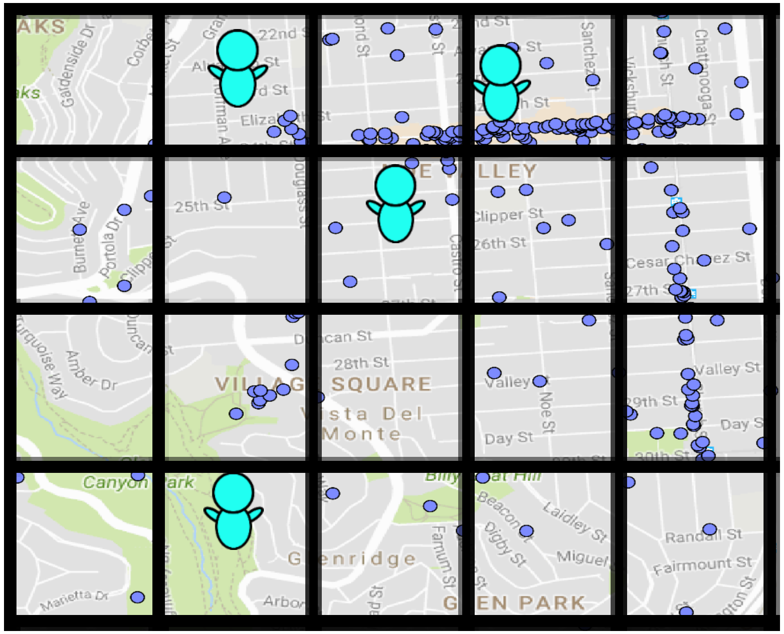
Sporadic

- Independent location reports.
- Adequate for infrequent usage (e.g., checking the weather)



Non-Sporadic





- Model how the user moves in the map.
- Typical computational constraints: discrete models.

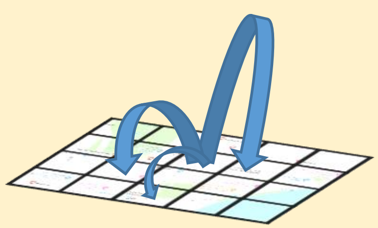
LPPM Design Notions: Metrics and Mobility Models

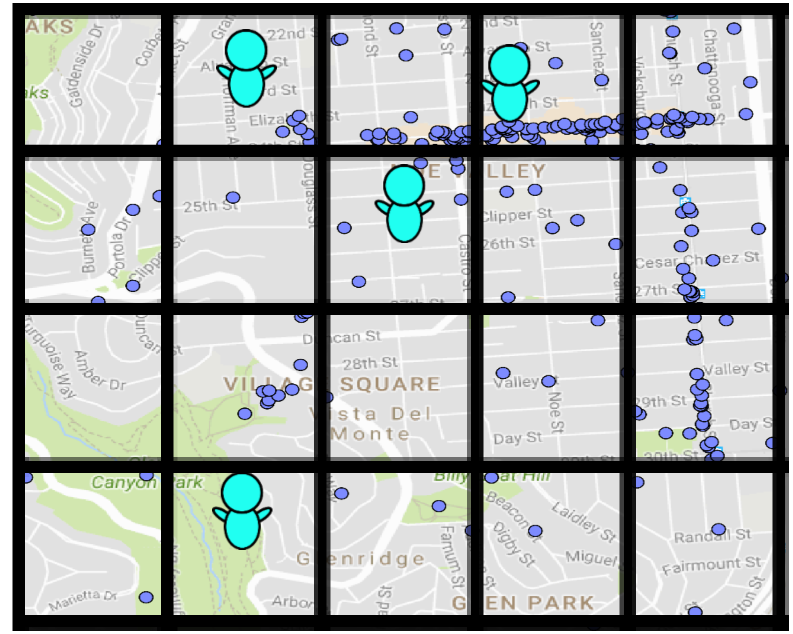
Sporadic

- Independent location reports.
- Adequate for infrequent usage (e.g., checking the weather)

Markov

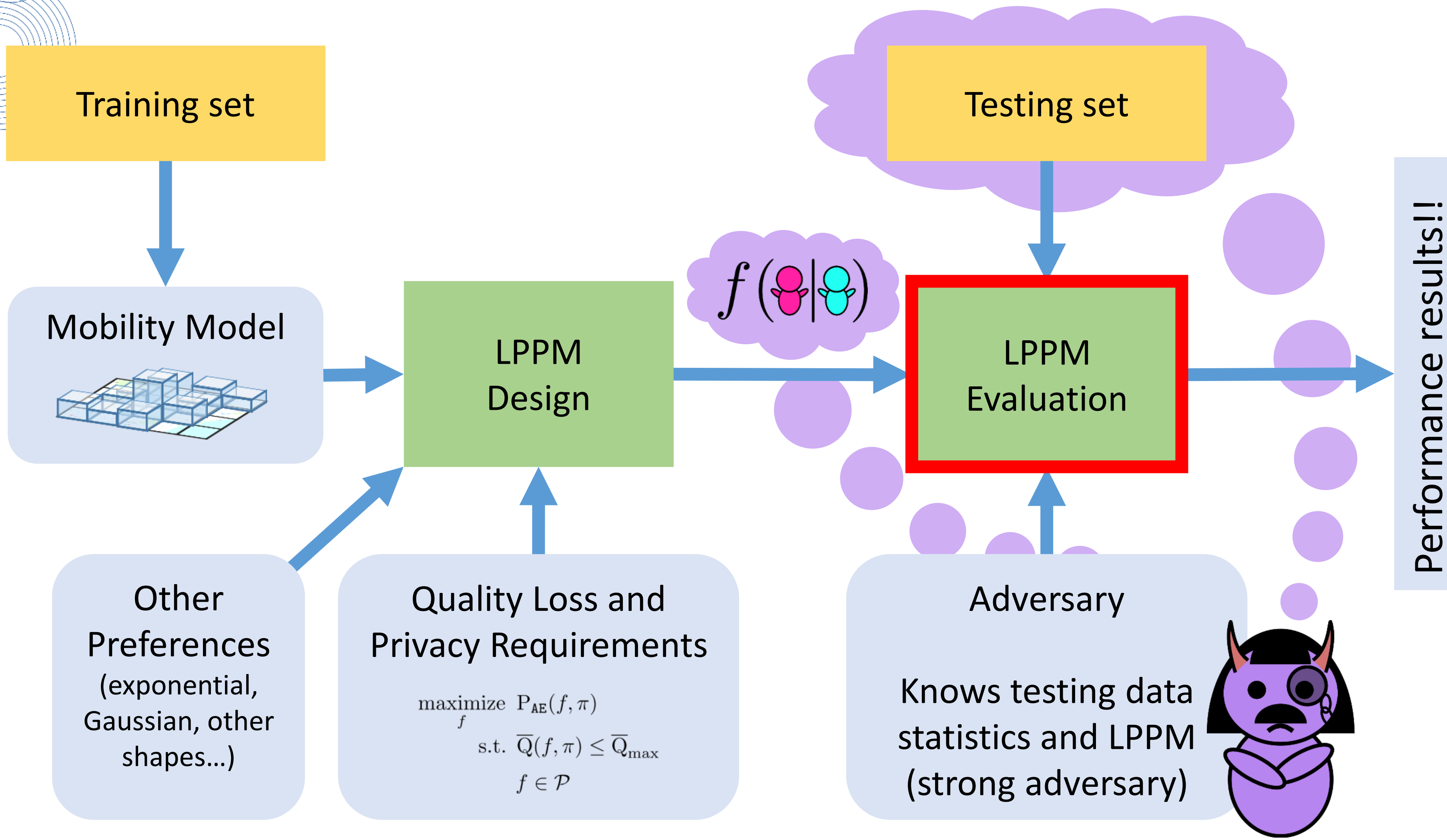
- Dependent locations
- Adequate for continuous usage (e.g., live location sharing)



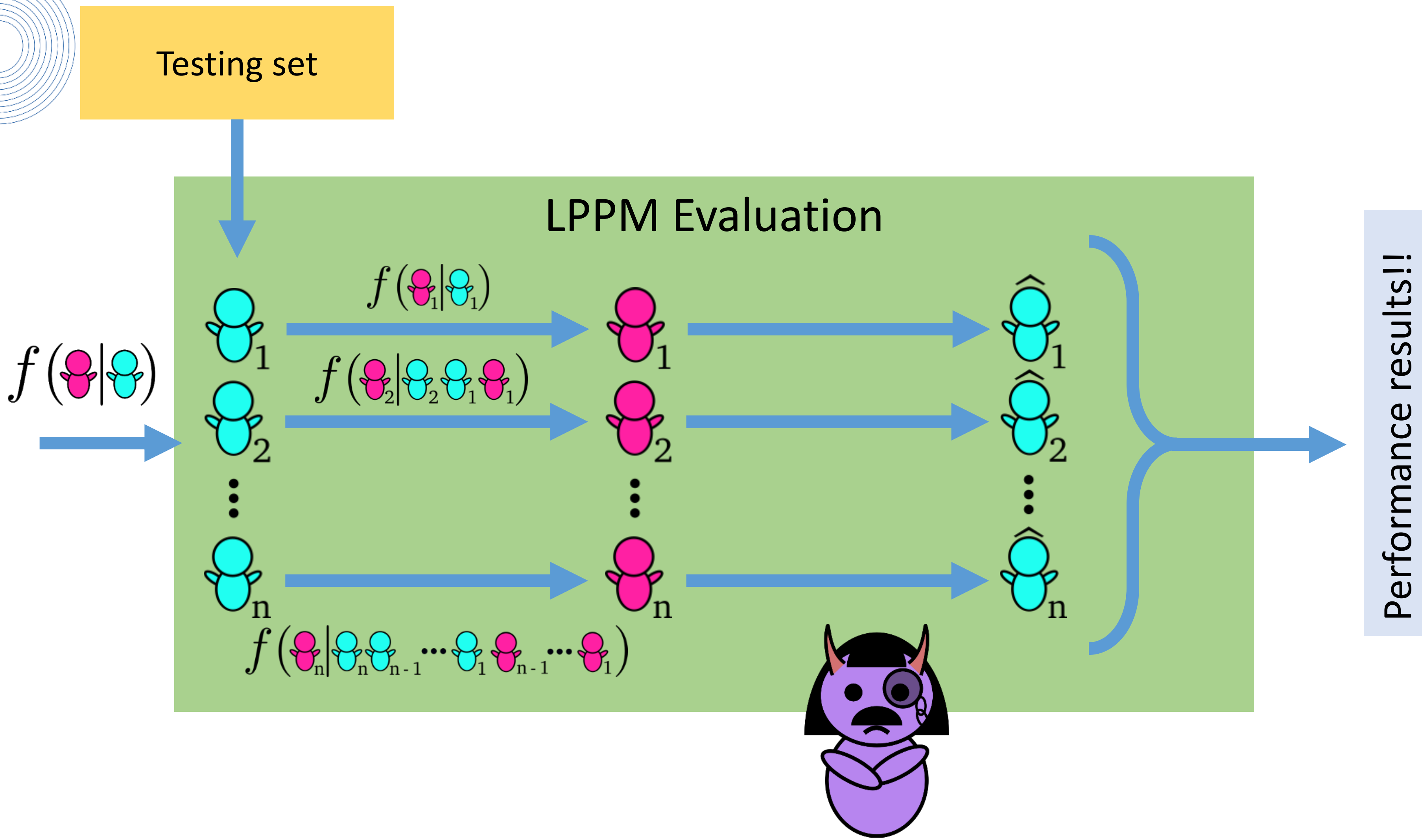


- Model how the user moves in the map.
- Typical computational constraints: discrete models.

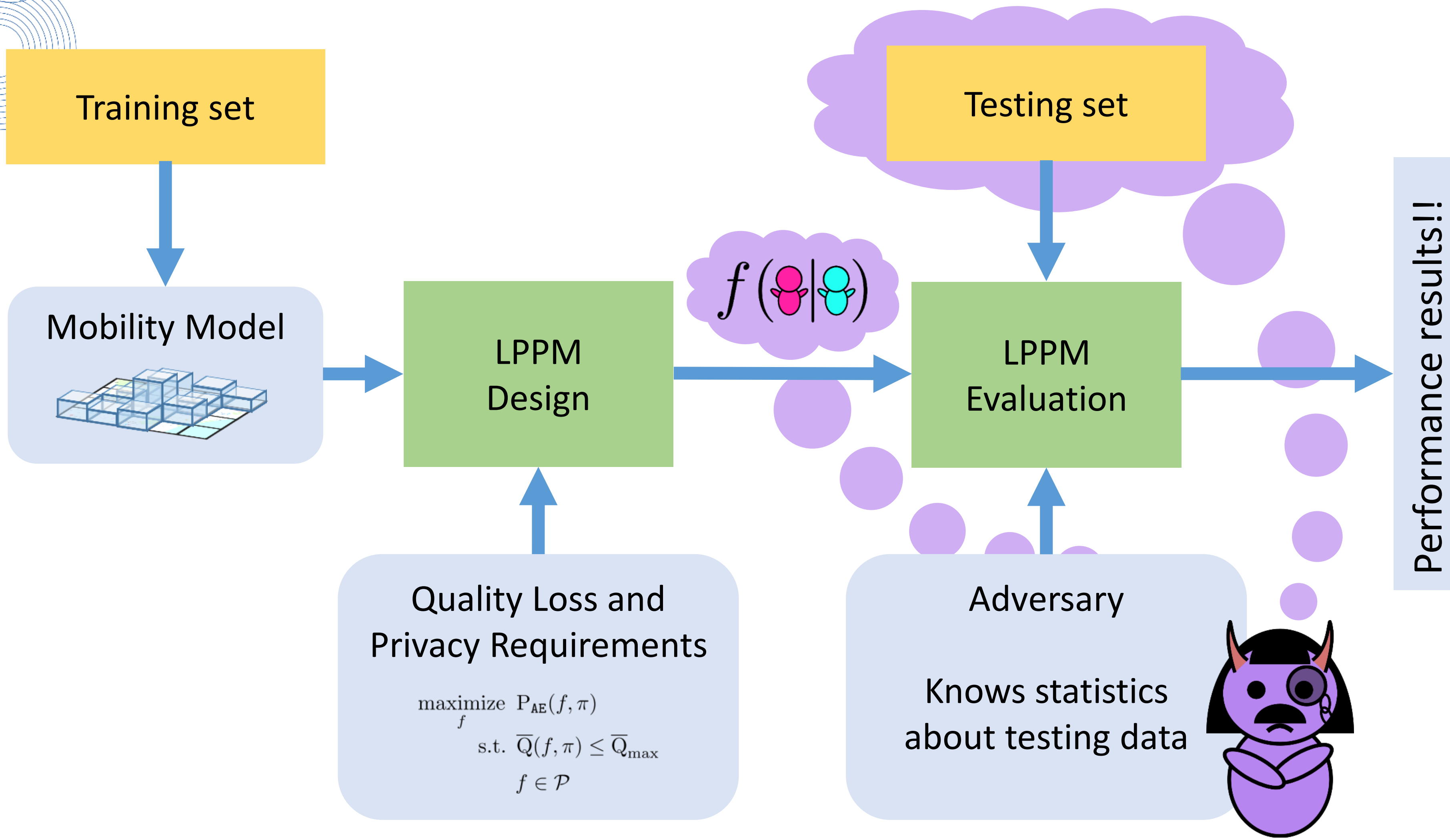
LPPM Design and Evaluation Framework



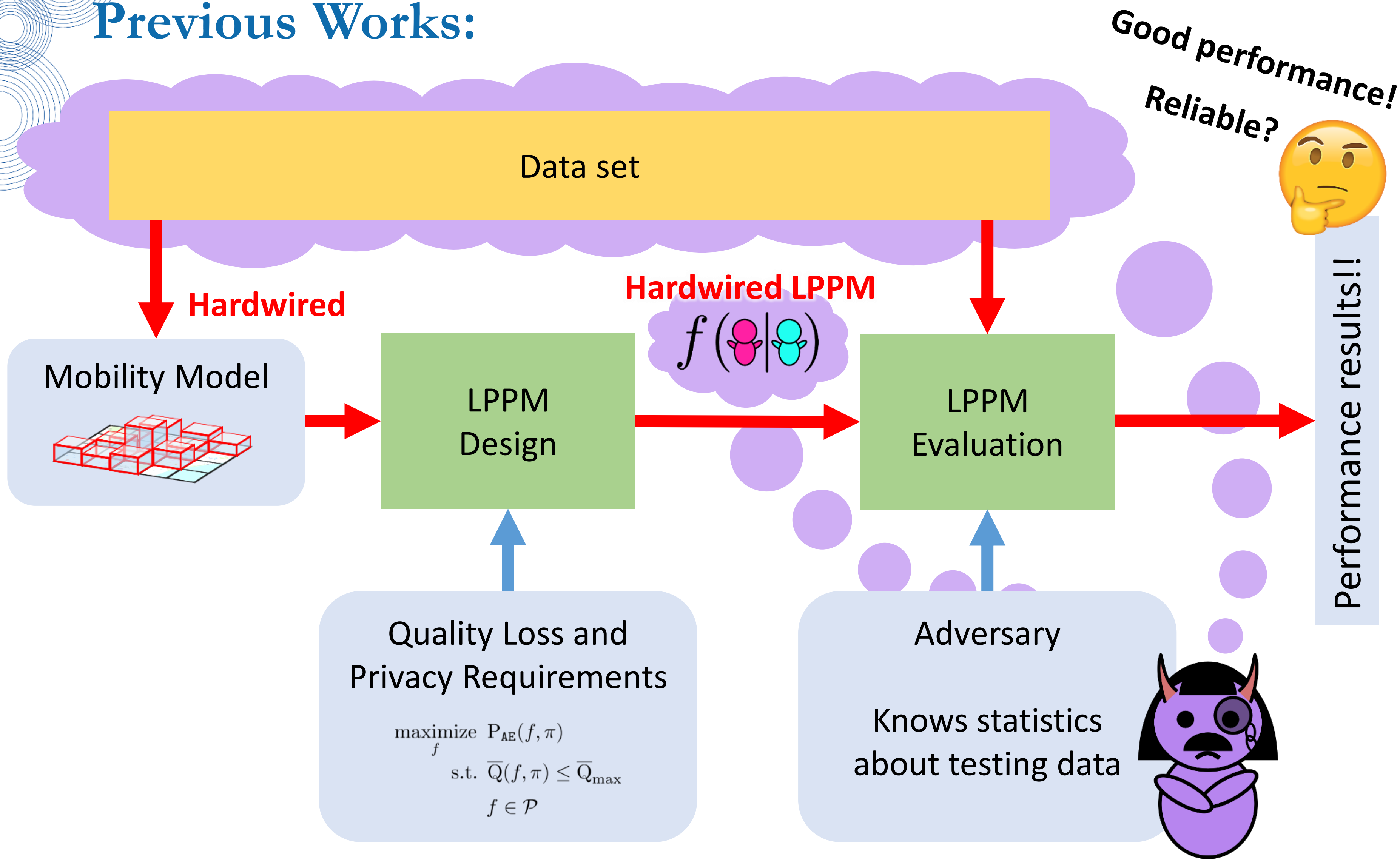
LPPM Design and Evaluation Framework



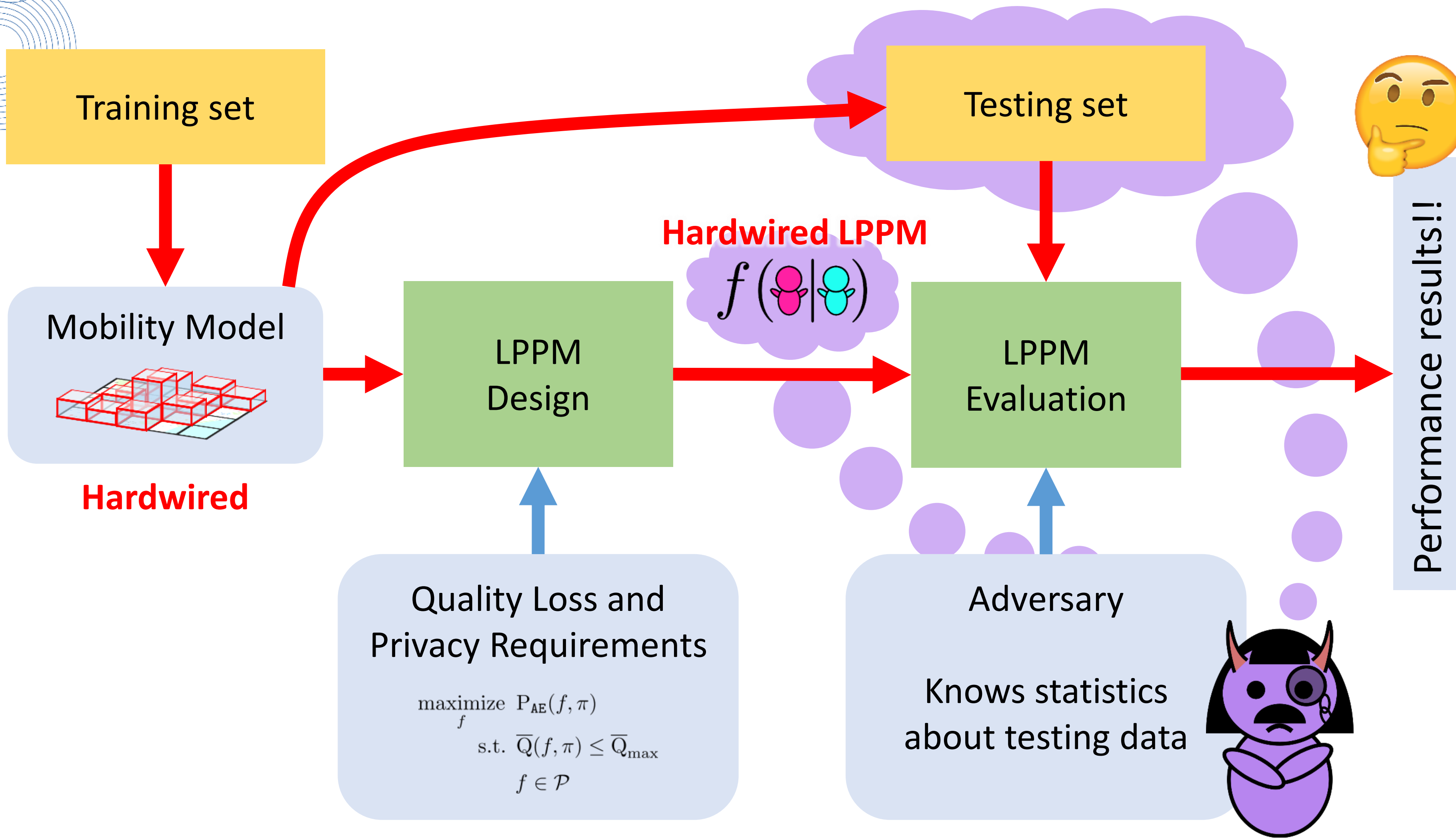
Previous Works:



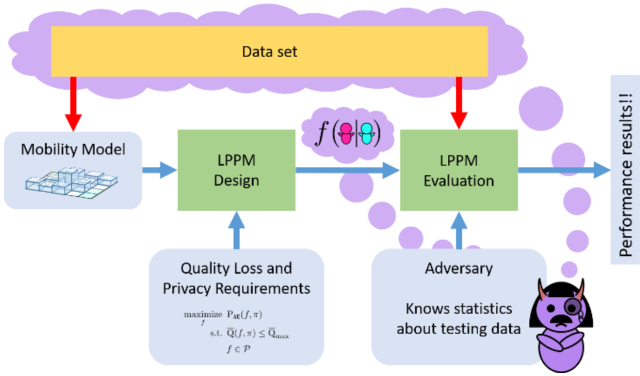
Previous Works:

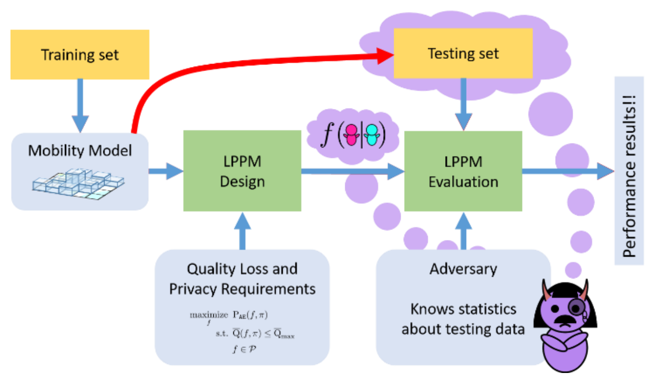


Previous Works:

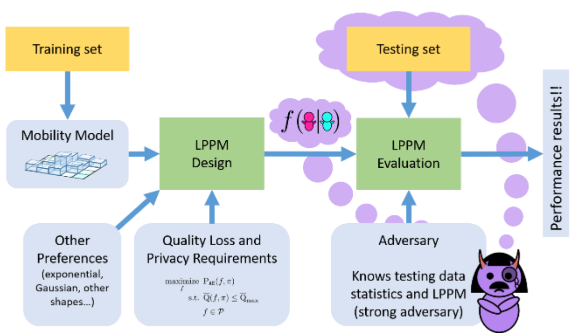


- In these frameworks, it makes sense to **hardwire** the training set into the LPPM:





- How do these LPPMs fare when we split training/testing data?



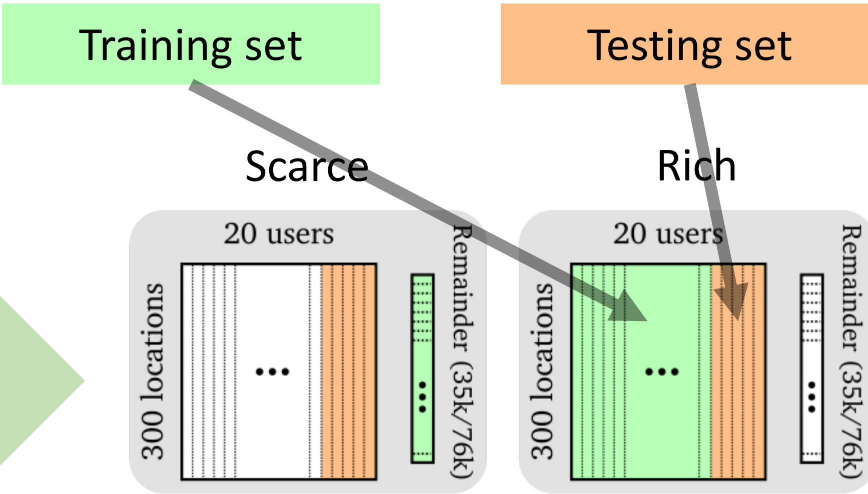
Experiment: let's see what would happen "in practice"

- Data gathering:

For sporadic mobility:

Gowalla
Brightkite

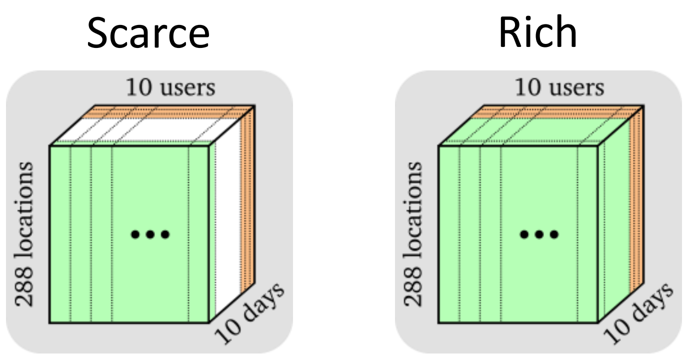
Pre-processing



Non-sporadic mobility:

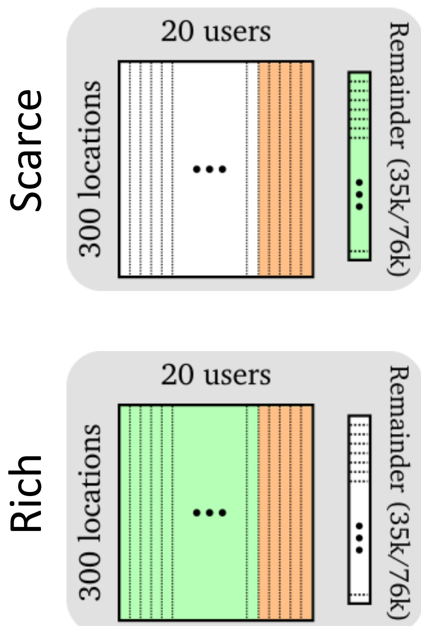
TaxiCab
(dense cab location reports for 30 days)

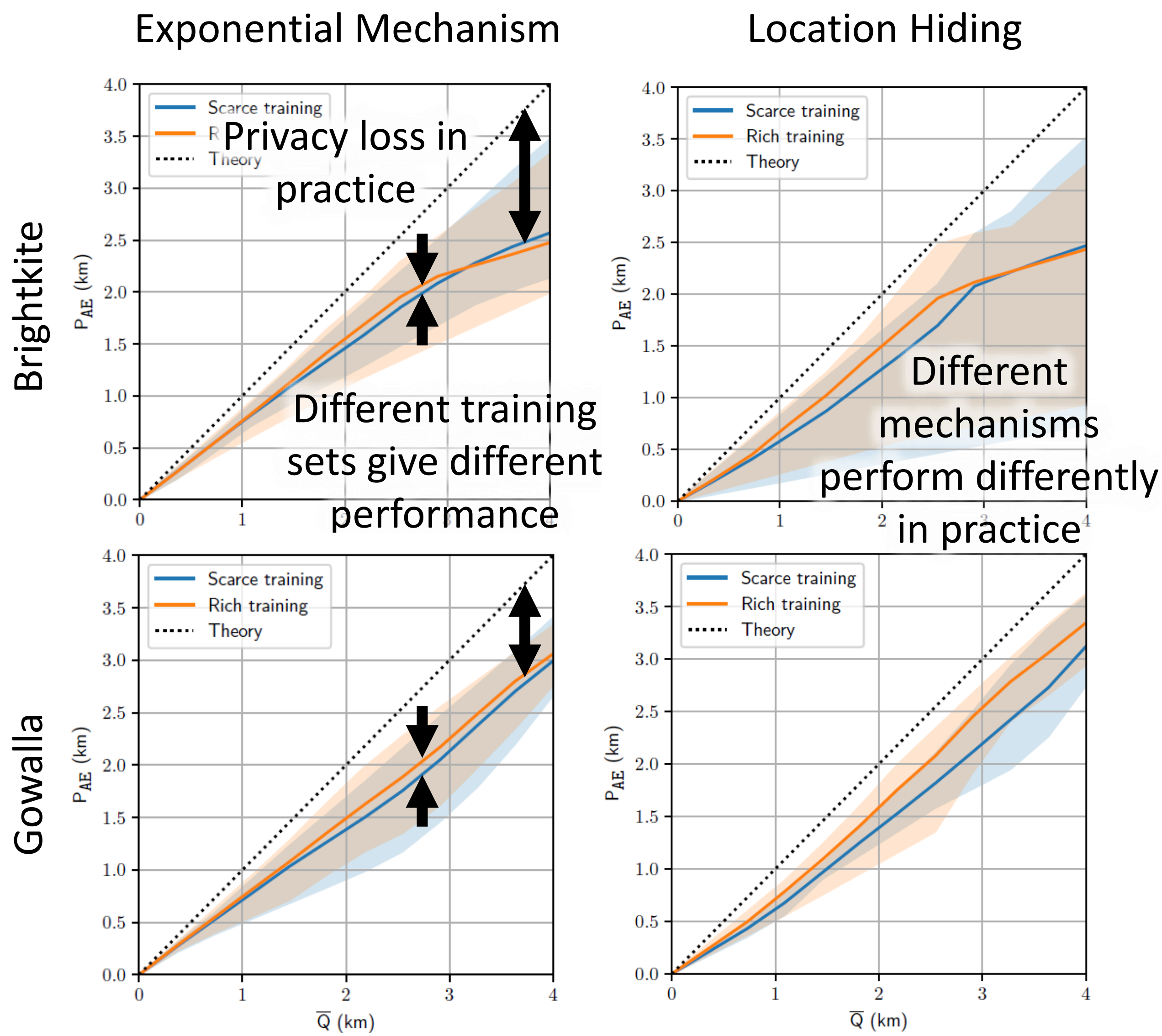
Pre-processing



Performance Results (sporadic case)

Datasets with sporadic reports (shuffled)

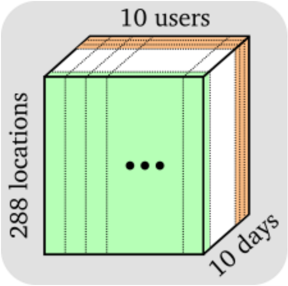




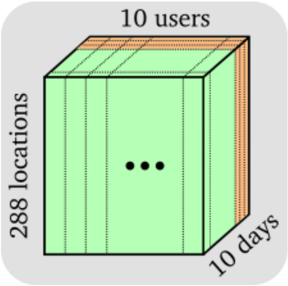
Performance Results (non-sporadic case)

TaxiCab
Dataset with
continuous reports

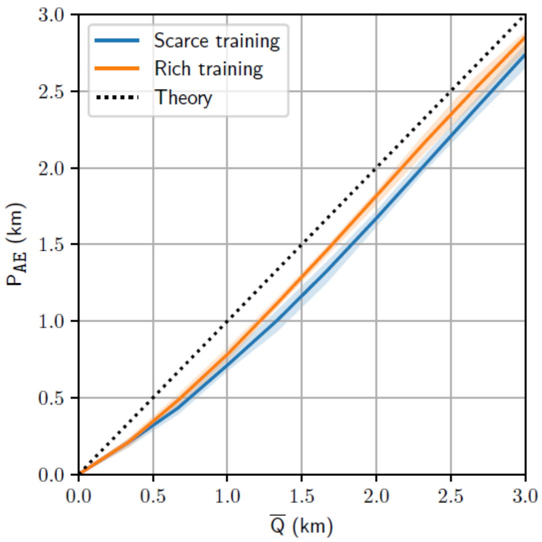
Scarce



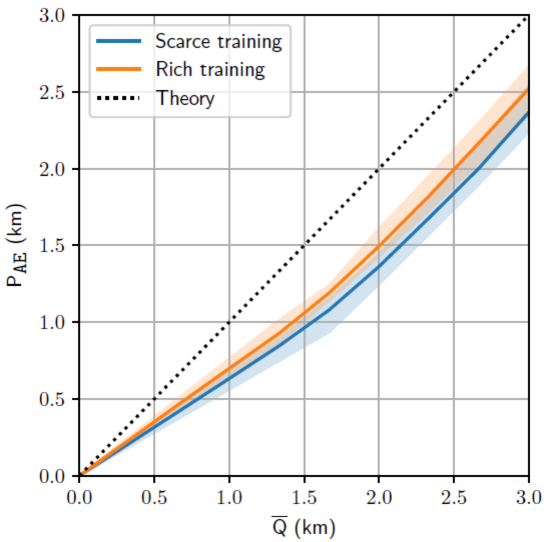
Rich



Exponential Mechanism



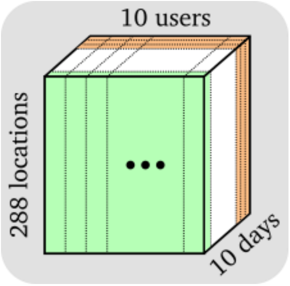
Location Hiding Mechanism



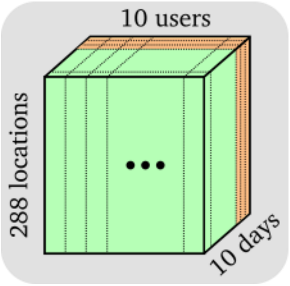
Performance Results (non-sporadic case)

TaxiCab
Dataset with
continuous reports

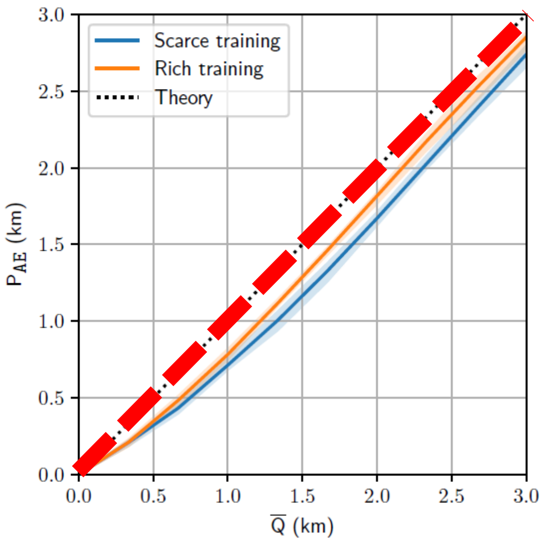
Scarce



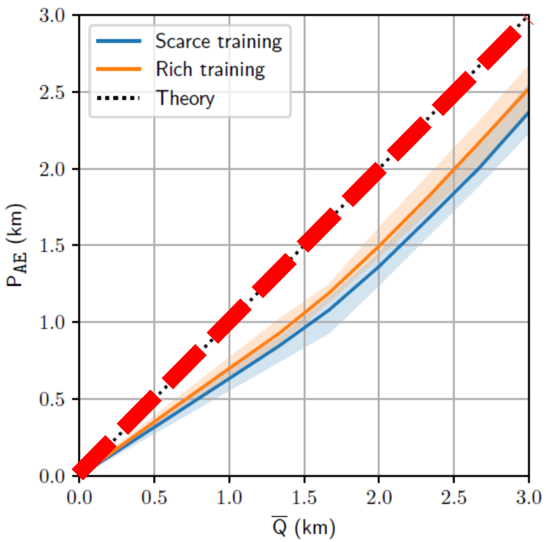
Rich



Exponential Mechanism



Location Hiding Mechanism

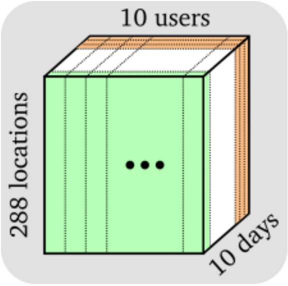


Same (optimal) performance in theory…

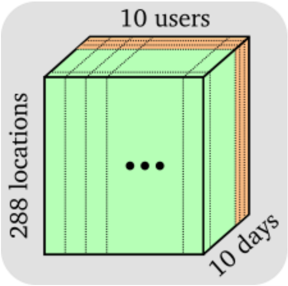
Performance Results (non-sporadic case)

TaxiCab
Dataset with
continuous reports

Scarce

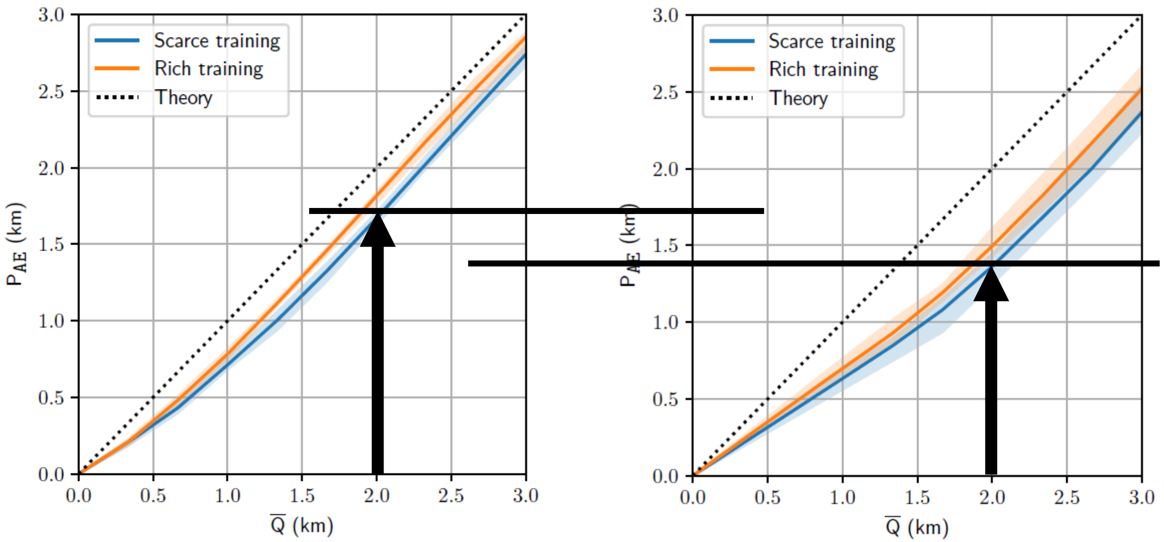


Rich



Exponential Mechanism

Location Hiding Mechanism



Same (optimal) performance in theory…

but different performance in practice

Let’s think about it…

- **Hardwired** LPPMs will be useful when user behavior (in practice) is captured by the training data.

Training set ≈ Testing set

- They will NOT perform well when:

Unknown Behavior

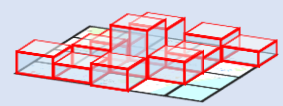
  - Insufficient data
  - Deprecated data
  - Non-representative data
  - Unexpected change in user behavior
  - …

- What can we do in all of these cases?

Training set

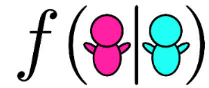
Hardwire

Mobility Model



LPPM Design

Hardwired LPPM



Let’s think about it…

- **Hardwired** LPPMs will be useful when user behavior (in practice) is captured by the training data.

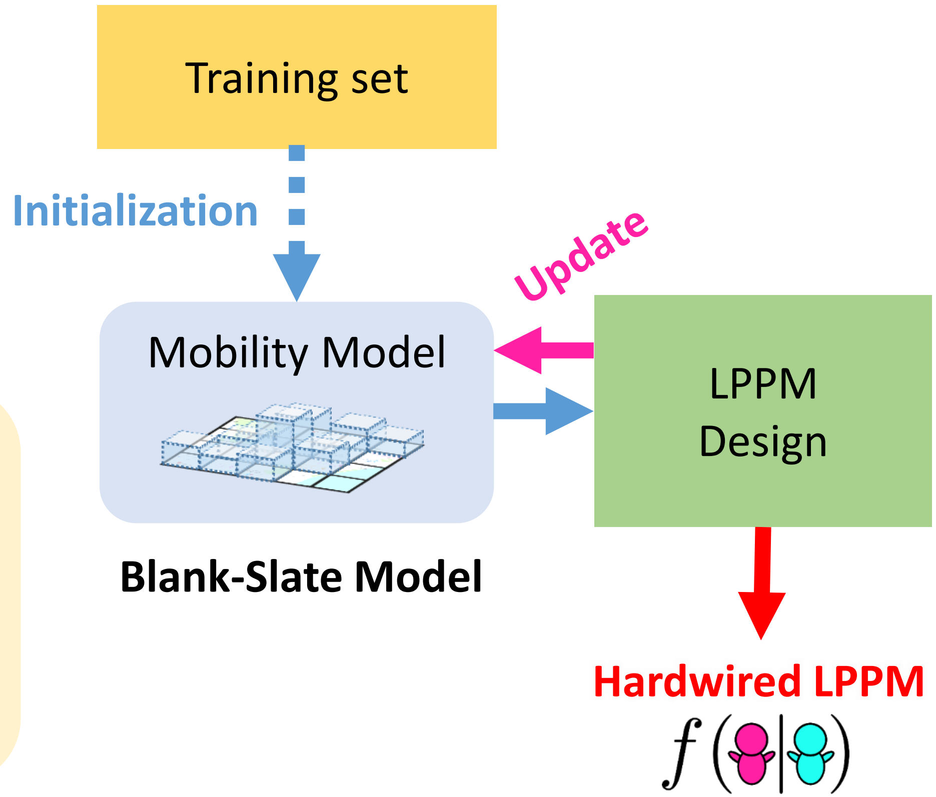
Training set ≈ Testing set

- They will NOT perform well when:

Unknown Behavior

  - Insufficient data
  - Deprecated data
  - Non-representative data
  - Unexpected change in user behavior
  - …

- What can we do in all of these cases?



Let’s think about it…

- **Hardwired** LPPMs will be useful when user behavior (in practice) is captured by the training data.

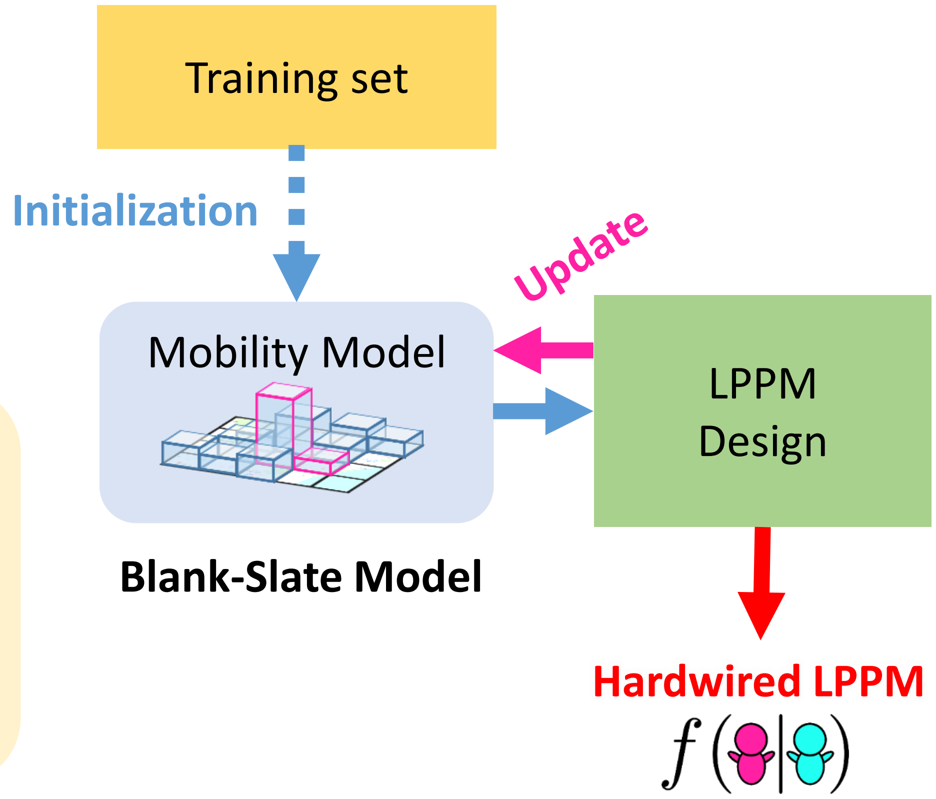
Training set ≈ Testing set

- They will NOT perform well when:

Unknown Behavior

  - Insufficient data
  - Deprecated data
  - Non-representative data
  - Unexpected change in user behavior
  - …

- What can we do in all of these cases?



Let's think about it…

- **Hardwired** LPPMs will be useful when user behavior (in practice) is captured by the training data.

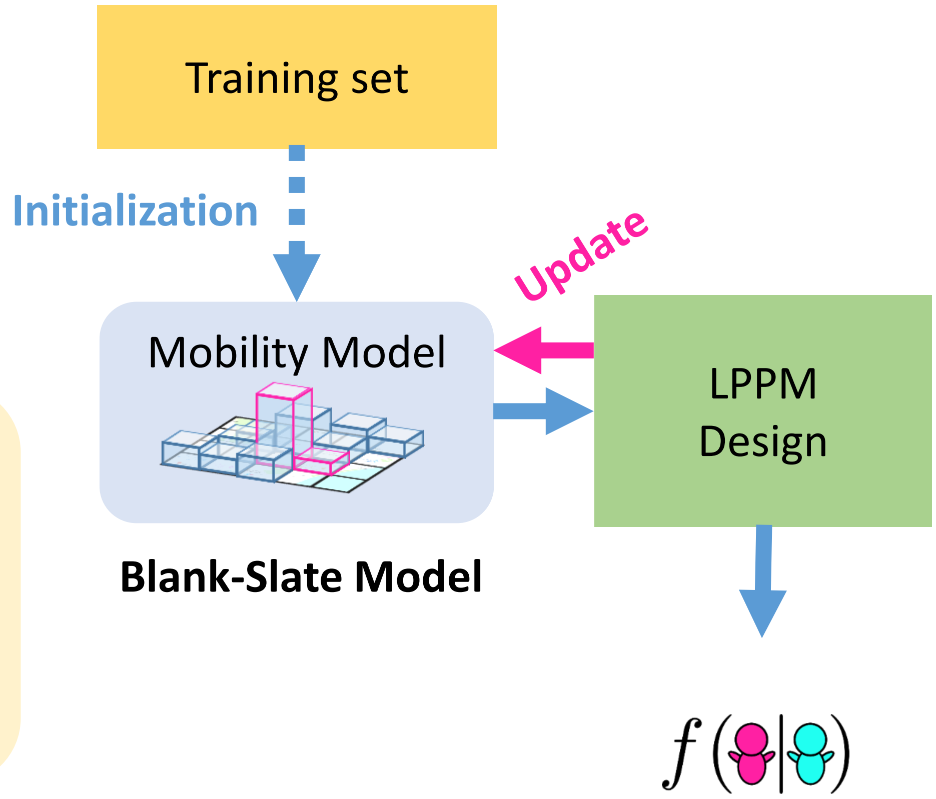
Training set ≈ Testing set

- They will NOT perform well when:

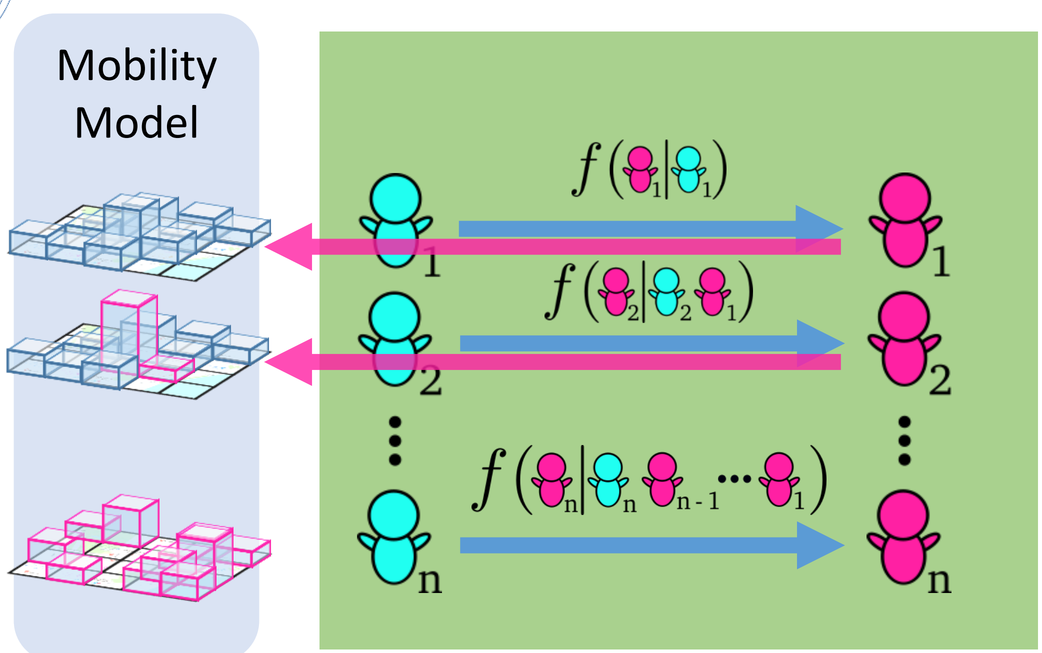
Unknown Behavior

  - Insufficient data
  - Deprecated data
  - Non-representative data
  - Unexpected change in user behavior
  - …

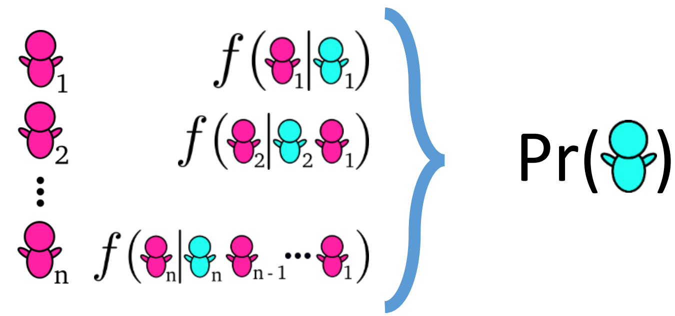
- What can we do in all of these cases?



Writing in the blank-slate using the reported locations



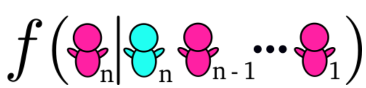
[In the paper]
MLE of the mobility profile
in **sporadic** models



Iterative algorithm

Profile Estimation-Based
(PEB) LPPMs

Result: an LPPM that can be written as:

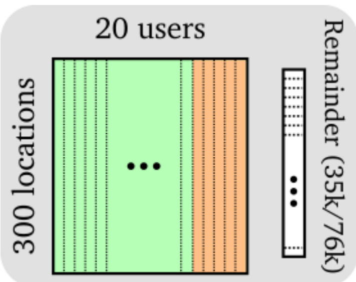


- We can evaluate them against a worst-case adversary.
- Will do better in sporadic settings.

Experimental Results. Sporadic Case

Datasets with sporadic reports (shuffled)

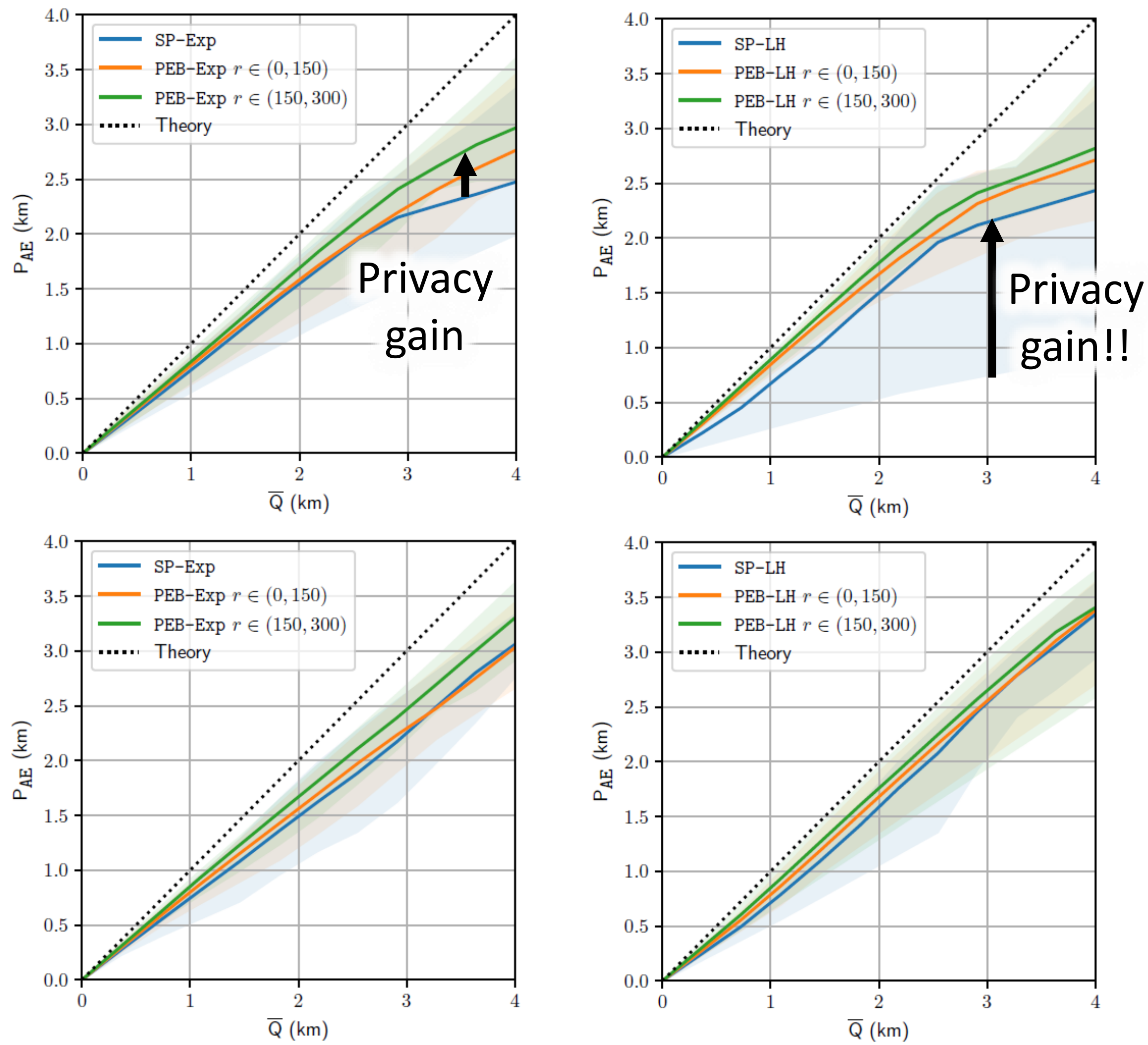
Rich



Exponential Mechanism | Location Hiding

Brightkite

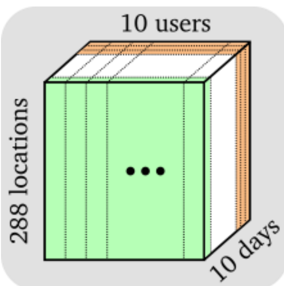
Gowalla



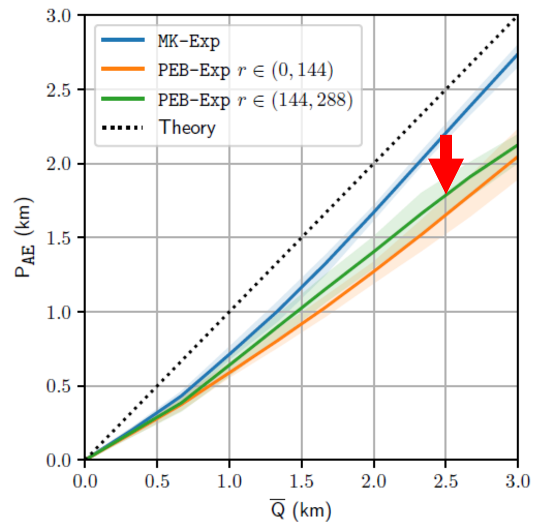
Experimental Results. Non-Sporadic Case.

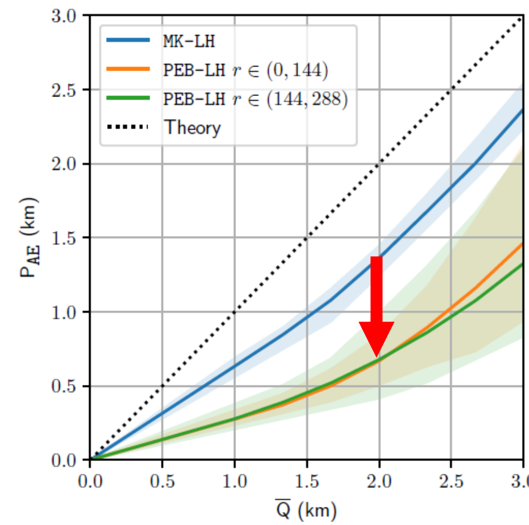
Dataset with
continuous reports

Scarce



TaxiCab

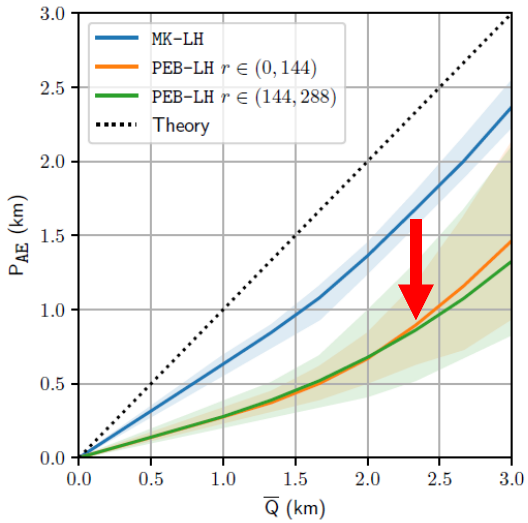




AtlantTIC Research Center for Information & Communication Technologies

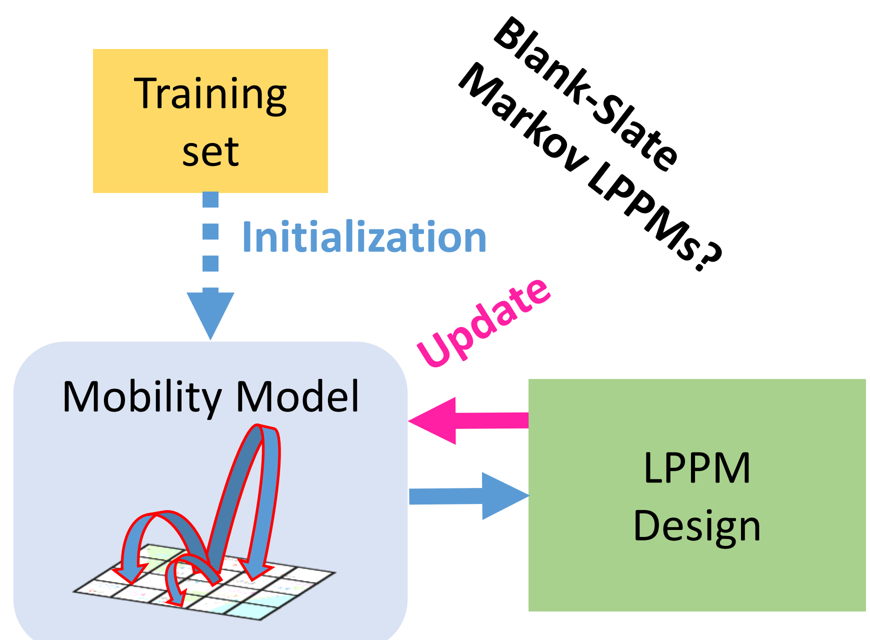
Universida de Vigo

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE



- However, current Markov LPPMs do not account for differences in train/test.

- Hardwired Markov-based LPPMs encode road restrictions.
- Sporadic PEB-LPPMs do not!
- This explains their difference in performance.



Sumary

To build PETs with strong privacy guarantees in practice, we have to embrace that training data cannot always capture user behavior.

Training set $\neq$ Testing set

- Current proposals **hardwire** training data into the LPPMs.
- We propose **blank-slate** models that improve the performance in sporadic scenarios.

Future Work

- Blank-slate Markov models
- Evaluate LPPMs with more data sets
- Develop other techniques to improve performance in practice…

Thank you!! simonoya@gts.uvigo.es

AtlantTIC Research Center for Information & Communication Technologies

UniversidadeVigo

GPSC

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE