

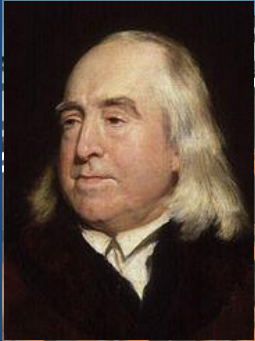
# Towards Intelligent Location-Privacy Preserving Mechanisms

Reza Shokri

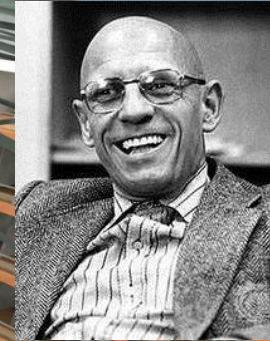
École polytechnique fédérale de Lausanne (EPFL)

April 2013

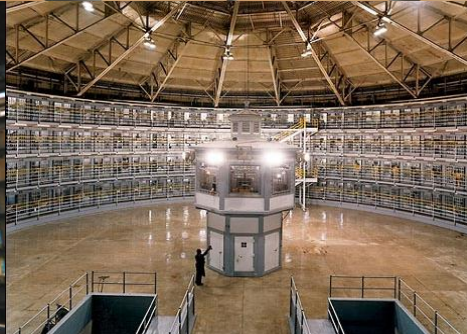
# Panopticon



Jeremy  
Bentham  
1748–1832



Michel  
Foucault  
1926–1984



# Panopticon: Age of Technology



**NOTHING IS PRIVATE. NOTHING IS SACRED.**

**POWERS**  
-POTENT AND UNMISSABLE!  
-Rolling Stone, TIME MAGAZINE  
-A NAIL-BITTER OF A THRILLER!  
-Los Angeles Times, Entertainment Weekly

**WINNER**  
-7th Annual Critics' Choice Awards  
-BEST PICTURE  
-BEST DIRECTOR  
-BEST ACTOR

**ACADEMY AWARDS**  
-BEST FOREIGN LANGUAGE FILM  
-Official Selection Berlin

**WINNER**  
-LONDON FILM FESTIVAL  
-BEST FILM  
-AUDIENCE AWARD

**THE LIVES OF OTHERS**  
A FILM BY FLORIAN HENCKEL VON DONNERSMARCK  
www.thelivesofothers.com www.dvds.com

1992  
March 18, 1992  
Gazette  
Page A-61

**Unless they're stopped, the Bells will know more about you than even the IRS.**

By keeping track of who you call, how often you call and how long your conversations last, the regional Bell telephone companies can create an alarmingly detailed profile of your life. Fortunately, they have never had any reason to do this. Until now.

The Bells are no longer satisfied with owning a monopoly over local phone lines. They want to own and control the information services that flow through those lines. By using their file on you, the Bells want to make you the target for all kinds of sales messages involving your business, your health, your finances, your family life and more.

Call an obstetrician and you could be deluged with information on diaper services and daycare centers. Call for stock quotes and they could try to sell you their stock selection service. Call a dating service and the next thing you know the Bells could try to arrange your social life.

In short, every time you picked up the phone, whether making a call or answering it, you'd be revealing something about yourself.

Something the Bells could take advantage of.

We need to stop this potential invasion of privacy. We need to keep the already thriving information services industry competitive and independent of the Bell monopoly.

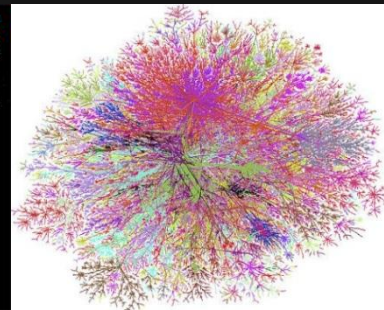
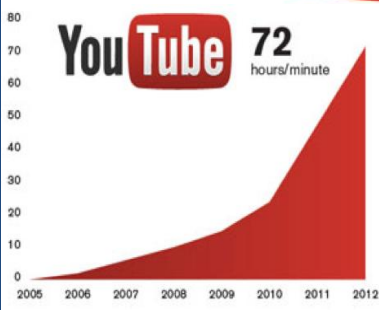
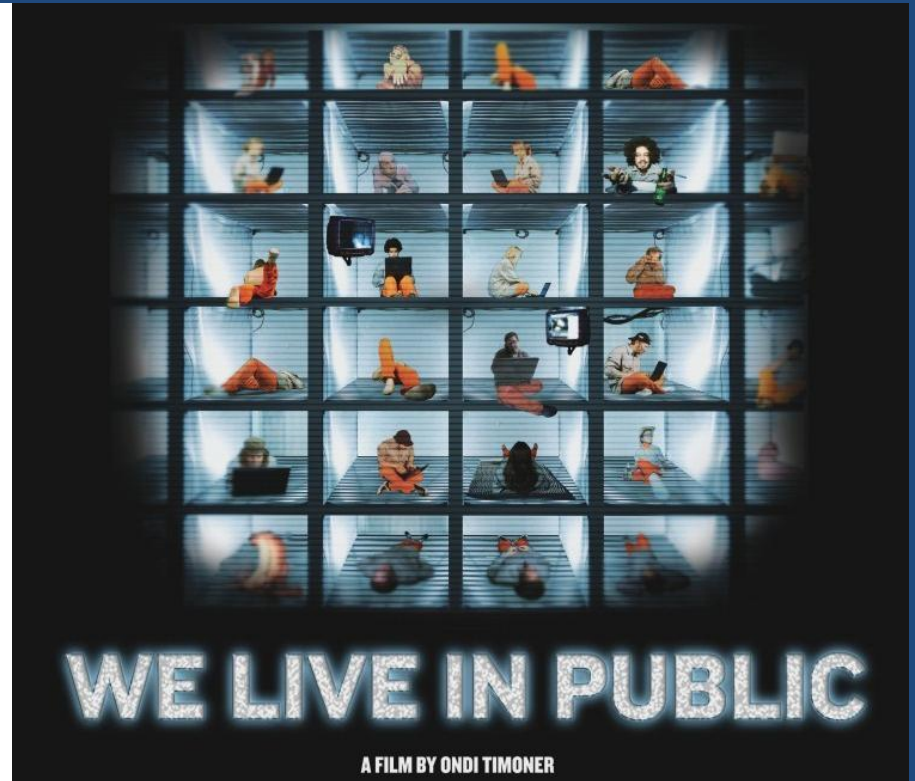
You can help by urging your U.S. Representative to support HR 3515. And by calling 1-800-54-PRIVACY.

Because if you remain silent now, everything you say later can, and just might, be used against you.

**Don't baby the Bells. Keep competition alive.**

American Newspaper Publishers Association • Consumer Federation of America • Dialog Information Services, Inc. Graphic Communications International Union • National Newspaper Association • Newspaper, Inc. The message describes Bell opportunities under current federal law. State requirements may vary.

# Panopticon: Age of Big Data



# If you are not paying for it: You Are The Product

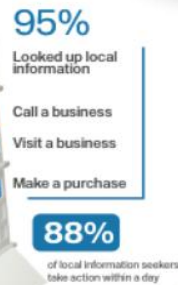
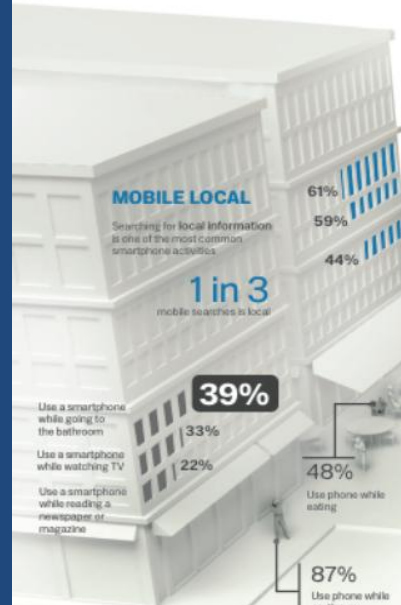


## Location-based Services

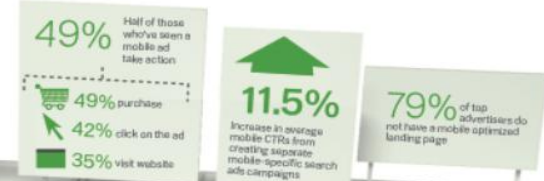


## The Constantly Connected Consumer

The everyday lives of US consumers are being transformed by the ability to constantly connect through always on, always available mobile devices. Understanding the impact of smartphones on consumer behavior is key to reaching today's on-the-go user. This growing consumer use of smartphones to search, shop and look for local information creates more opportunities for marketers to connect with their customers.



## Behavioral Advertising



Source: US Mobile Smartphone Consumer Study, Ipsos & Google, 2010; Google Internal Data

## MOBILE BEHAVIOR

Smartphones have become such an indispensable part of our daily lives



## MOBILE SHOPPING

Smartphones have become the ultimate shopping companion

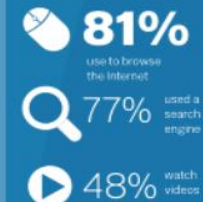


## MOBILE SEARCH



## SMARTPHONE USAGE

In a typical week



## WHO USES A SMARTPHONE?



Studies show **smartphone usage** will increase significantly this year

# You Are The Target

## User-Data Requests By Governments



POWERED BY  
Google

Map data ©2013 MapLink, Tele Atlas - Terms of Use

2

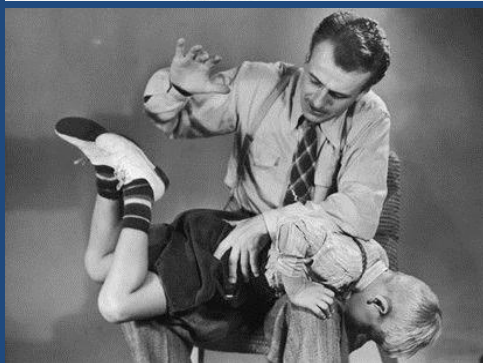
# The Timeline of Panopticon



Global and Ubiquitous



Personal

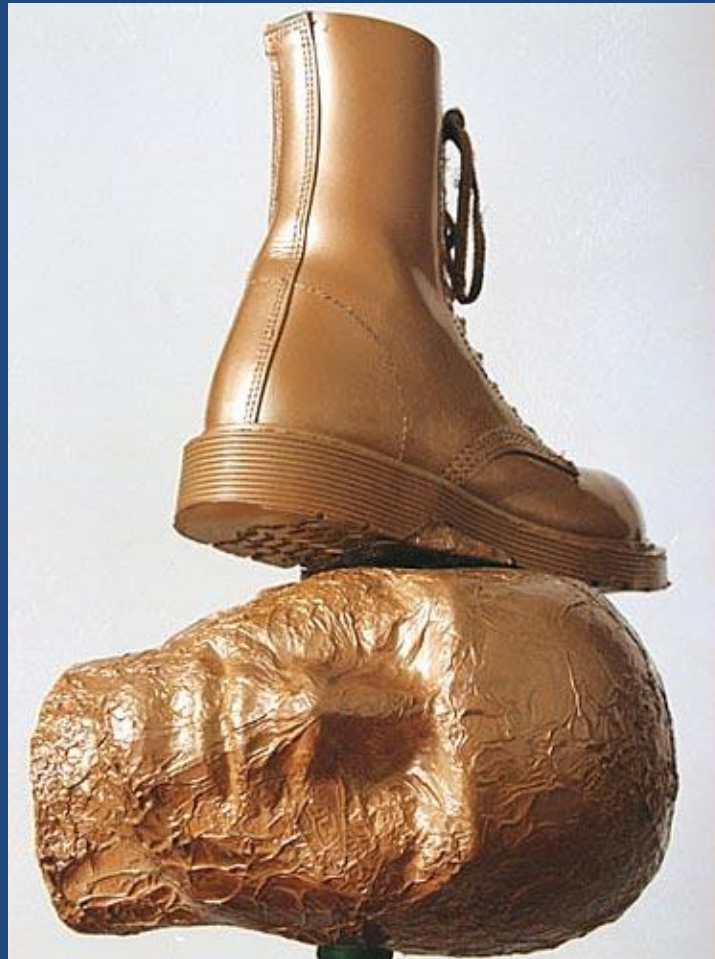


Reward-Oriented



# Lack of Privacy

- Imbalance of Power
- Influence

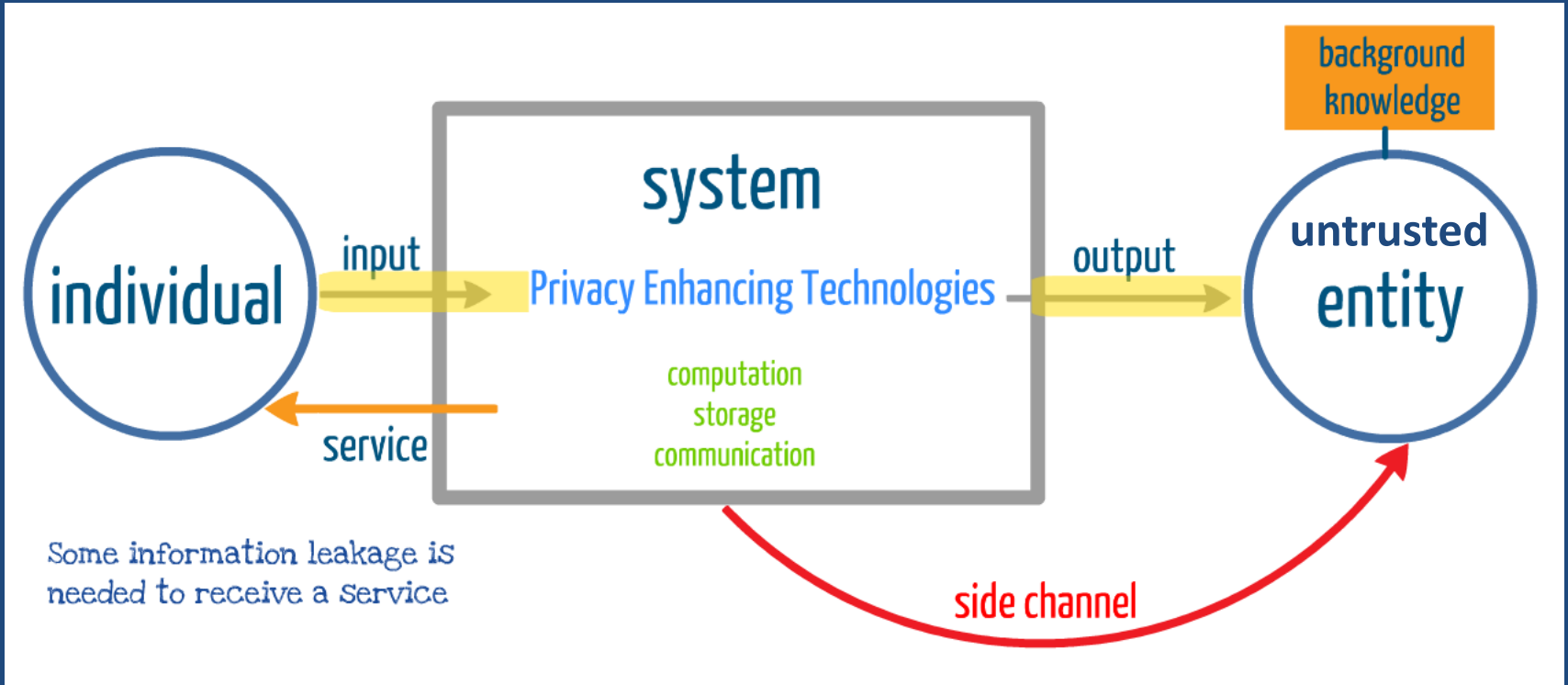




# Protecting Privacy

- Behavioral
- Legal
- Computational

# Computational Privacy



Quantitative information flow, Differential privacy, Bayesian analysis of Mix networks, ...

# Computational Privacy

## For Location-tagged Data Sharing

- **Quantifying Privacy**
  - Help individuals to accurately estimate their privacy risks
- **Protecting Privacy (in existing systems)**
  - Help individuals to find effective obfuscation mechanisms
- **Intelligent Tools and Technologies**
- **Focus of this talk: Location-Privacy of Mobile Users**



# Location-based Services

Exposed Location Trace  
[Who, When, Where]

Protect Privacy  
Distort Information

Anonymize



Use Pseudonyms

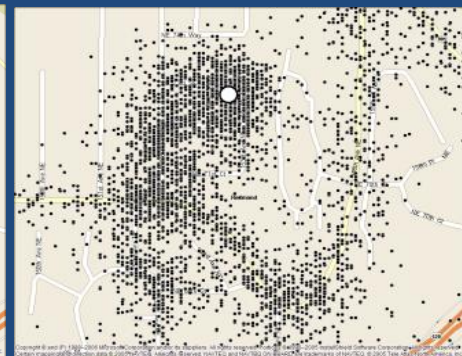
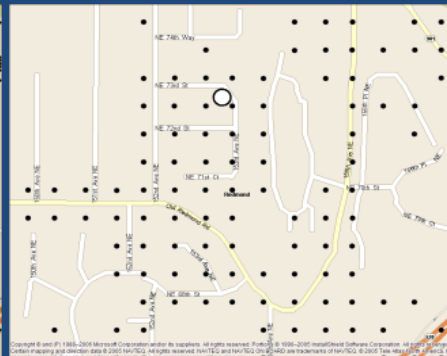
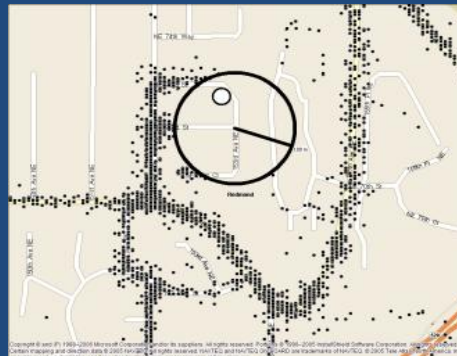
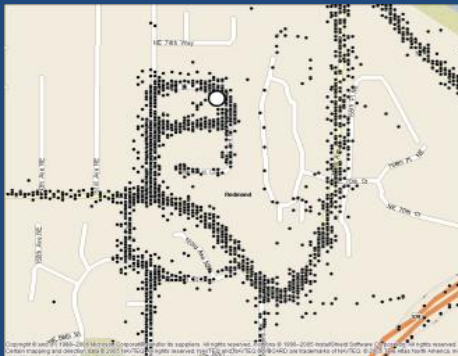
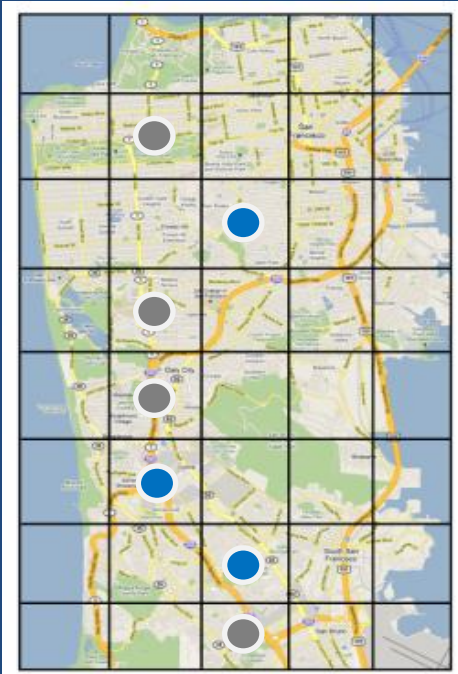
Obfuscate

Original Location

Hide Around Home

Low Precision

Low Accuracy



# Issues/Challenges

## Quantification

- Evaluate/Compare Various Protection Mechanisms
- Find the Right Metric for Quantifying Location Privacy
- Incorporate User's Data Model (e.g., Mobility Model)

## Protection

- Design Effective yet Useful Protection Mechanisms
- Respect User's Service Quality Requirements
- Incorporate User's Privacy Requirements/Sensitivities

# Quantifying Location Privacy

## Privacy Meter

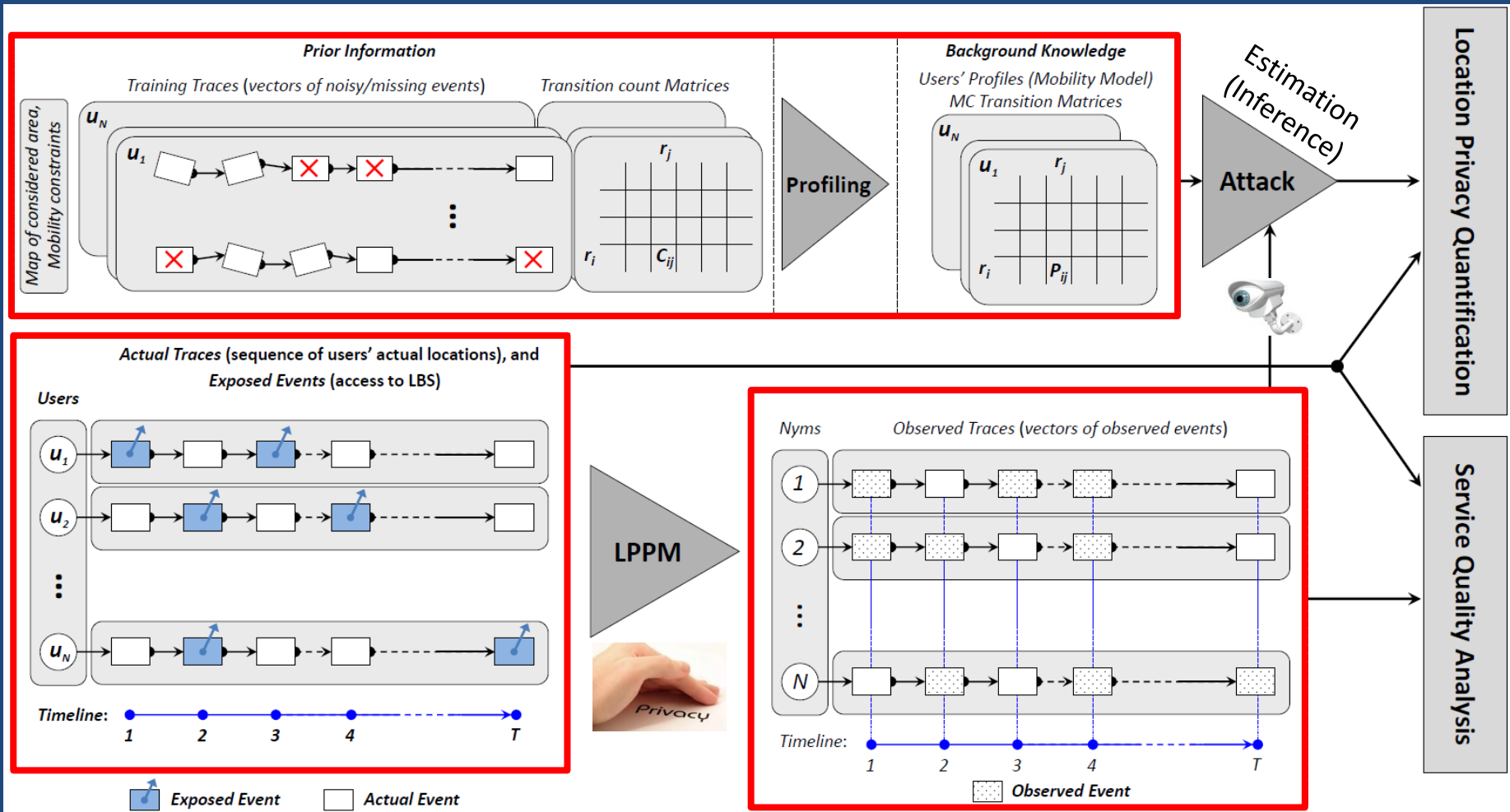
IEEE S&P (Oakland) 2011. R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux.

PETS 2011. R. Shokri, G. Theodorakopoulos, G. Danezis, J.-P. Hubaux, and J.-Y. Le Boudec.

# Approach

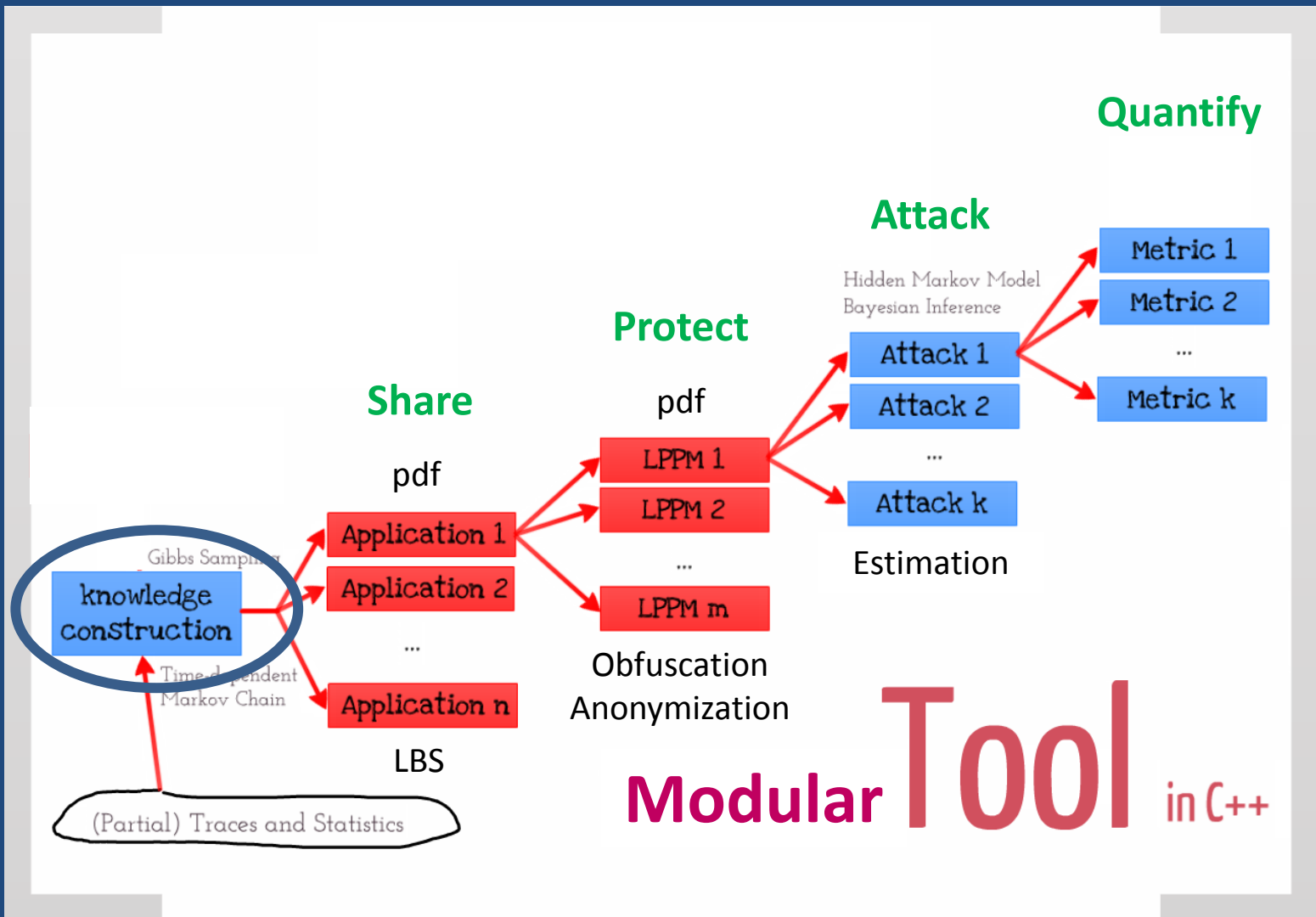
- Design a probabilistic framework
  - Formal definition of users/LBSs/defenses
- Turn the evaluation of a Location-Privacy Protection Mechanism (LPPM) to an estimation problem
- Throw attacks at the LPPM: Bayesian Inference
  - Metric: Estimation Error
- Design and implement a software tool
  - Location-Privacy Meter (LPM)

# A Probabilistic Framework Location Privacy





# Location-Privacy Meter (LPM)



# Knowledge Construction: User Profiling

**Mobility: Markov Chain**  $p_{r,r'}(u) = \mathbb{P}_R\{\mathbf{A}_u^{t+1} = r' \mid \mathbf{A}_u^t = r\}$

**Estimating transition probabilities given available traces and mobility constraints**

$$\mathbb{E}\{\hat{\mathbf{p}} \mid \mathbf{y}, \mathbf{c}\}$$

$$\mathbb{P}_R\{\hat{\mathbf{p}} \mid \mathbf{y}, \mathbf{c}\} = \sum_{\hat{\mathbf{y}}} \mathbb{P}_R\{\hat{\mathbf{p}}, \hat{\mathbf{y}} \mid \mathbf{y}, \mathbf{c}\}$$

**Gibbs Sampling**

$$\begin{aligned} (\hat{\mathbf{p}}^{\{l\}}, \hat{\mathbf{y}}^{\{l\}}) &\sim (\mathbb{P}_R\{\hat{\mathbf{p}} \mid \hat{\mathbf{y}}^{\{l-1\}}, \mathbf{y}, \mathbf{c}\}, \hat{\mathbf{y}}^{\{l-1\}}) \\ (\hat{\mathbf{p}}^{\{l\}}, \hat{\mathbf{y}}^{\{l\}}) &\sim (\hat{\mathbf{p}}^{\{l\}}, \mathbb{P}_R\{\hat{\mathbf{y}} \mid \hat{\mathbf{p}}^{\{l\}}, \mathbf{y}, \mathbf{c}\}) \end{aligned}$$

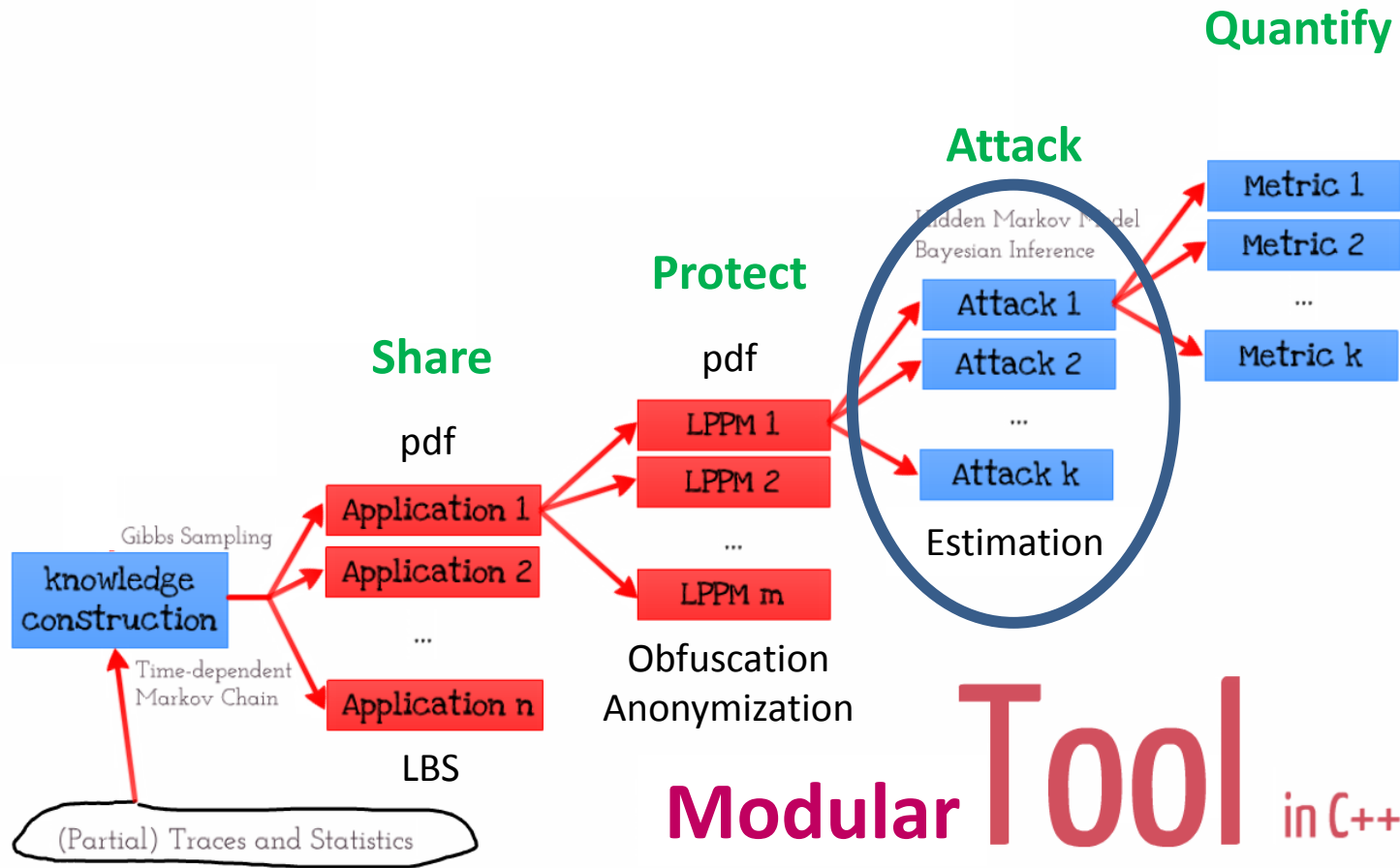
$$\hat{p}_{r,r'} = \frac{1}{L} \sum_l \hat{p}_{r,r'}^{\{l\}}, \forall r, r'$$

r: location (region)

y: location trace (potentially incomplete)

c: mobility constraints matrix (of 0/1)

# Location-Privacy Meter (LPM)



# Quantifying Location Privacy

**De-anonymization (re-identification)** Which observed trace is Alice's?

$$\sigma^* = \arg \max_{\sigma} \Pr\{\Sigma = \sigma \mid \mathbf{O} = \mathbf{o}\}$$

**Localization** Where was Alice yesterday at 10am?

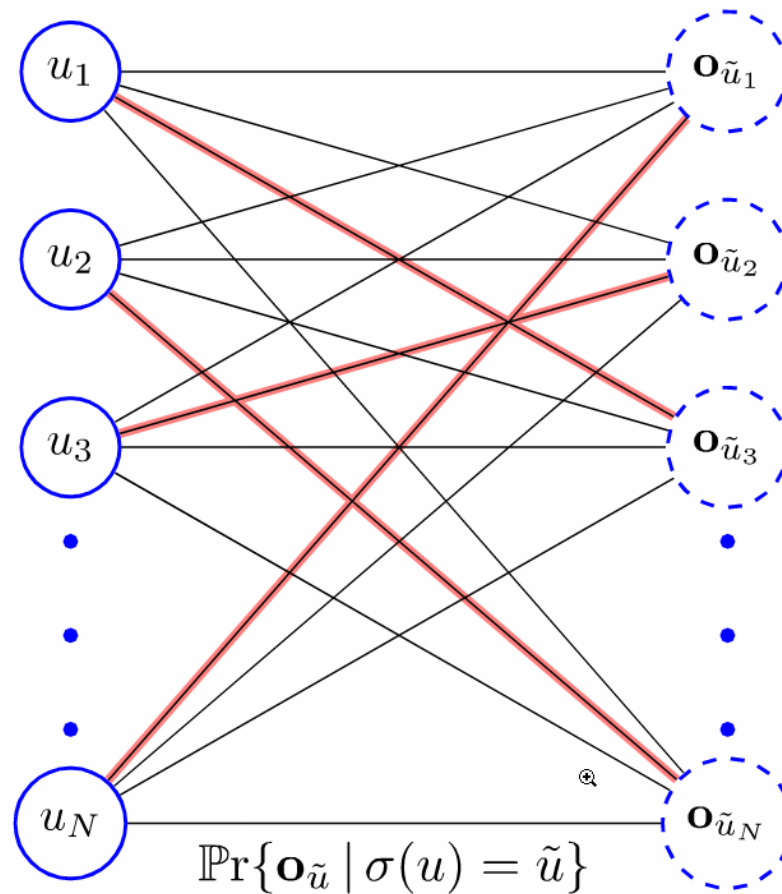
$$\Pr\{\mathbf{A}_u^t = r \mid \mathbf{o}_{\tilde{u}}, \sigma^*(u) = \tilde{u}\}$$

<i>a</i>	<i>actual</i>
<i>o</i>	<i>observed</i>
<i>σ</i>	<i>pseudonym</i>

**Privacy of users: expected estimation error of adversary in his inference attacks**

# De-Anonymization Attack

## Maximum Weight Assignment



# De-Anonymization Attack

$$\Pr\{\mathbf{o}_{\tilde{u}} \mid \sigma(u) = \tilde{u}\} = \sum_{\mathbf{a}_u} \Pr\{\mathbf{o}_{\tilde{u}} \mid \mathbf{a}_u, \sigma(u) = \tilde{u}\} \cdot \Pr\{\mathbf{a}_u\}$$

exponential complexity

**forward variables**

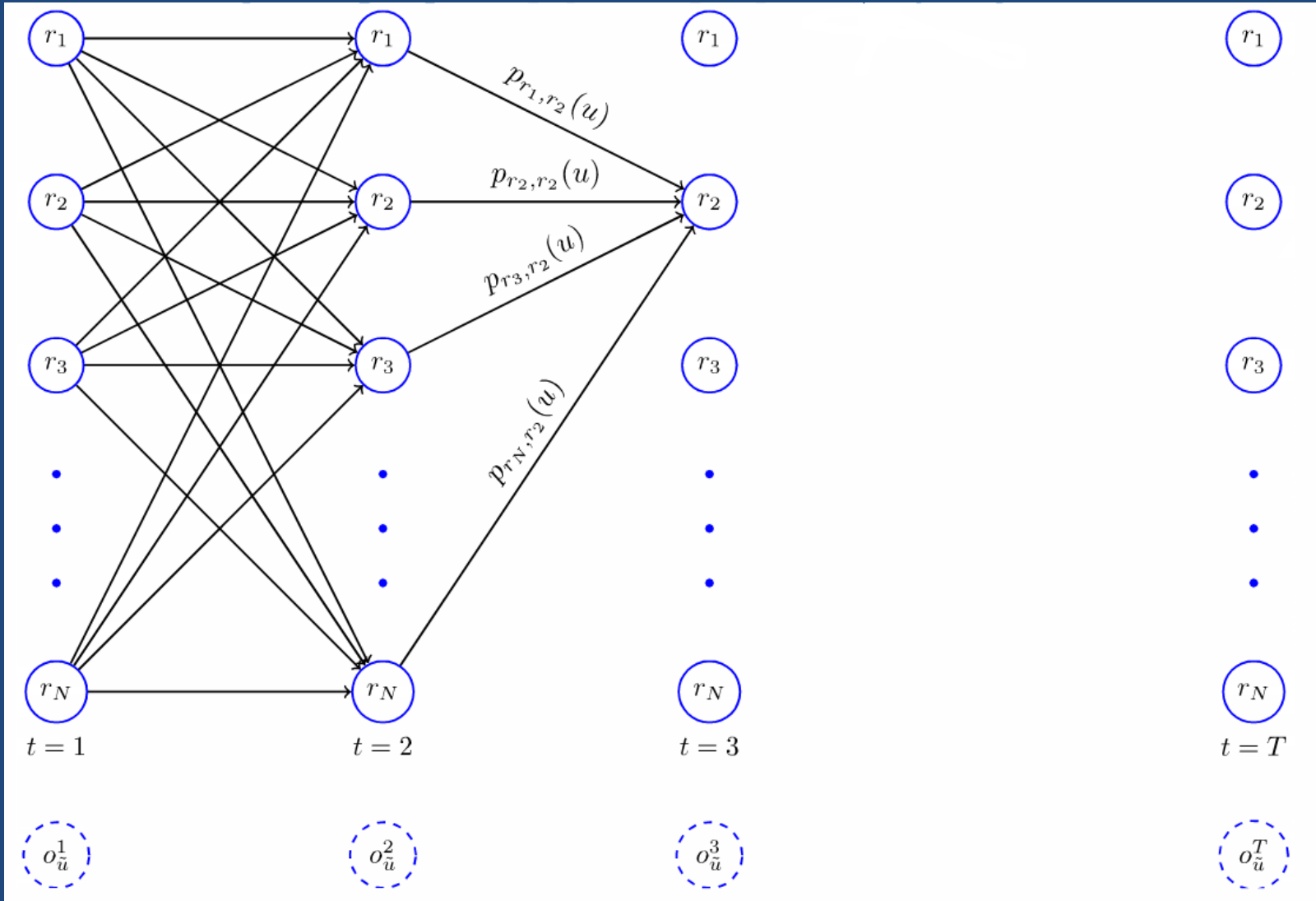
$$\alpha_r^{u, \tilde{u}}(t) = \Pr\{\mathbf{A}_u^t = r, \mathbf{o}_{\tilde{u}}^{1:t} \mid \sigma(u) = \tilde{u}\}$$

**likelihood**

$$\Pr\{\mathbf{o}_{\tilde{u}} \mid \sigma(u) = \tilde{u}\} = \sum_{r \in \mathcal{R}} \Pr\{\mathbf{A}_u^T = r, \mathbf{o}_{\tilde{u}} \mid \sigma(u) = \tilde{u}\} = \sum_{r \in \mathcal{R}} \alpha_r^{u, \tilde{u}}(T).$$

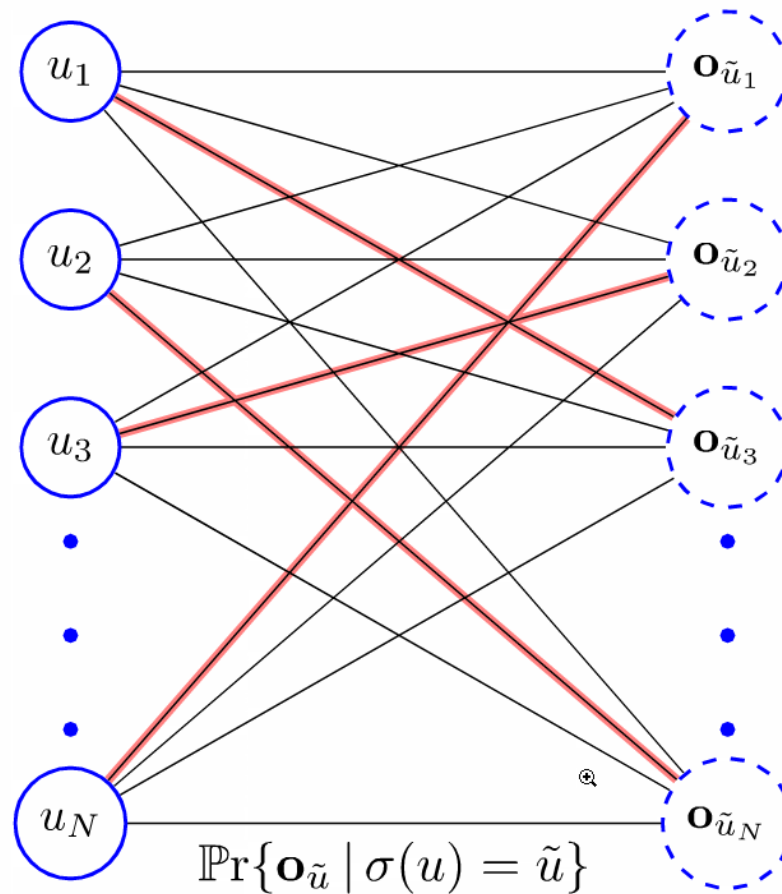
Hidden Markov Models

# De-Anonymization Attack



# De-Anonymization Attack

## Maximum Weight Assignment

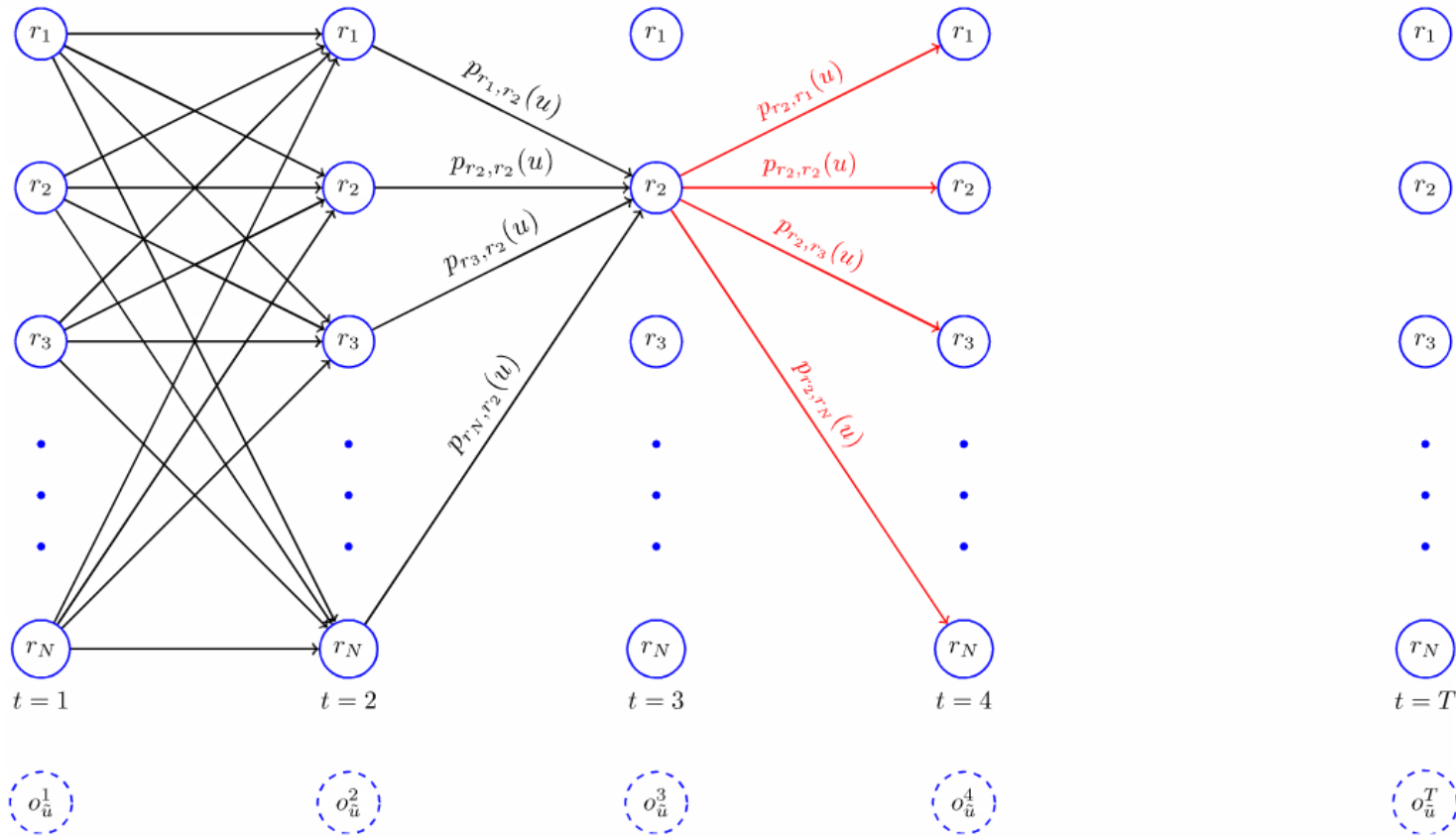




# Localization Attack

$$\mathbb{P}\Pr\{\mathbf{A}_u^t = r \mid \mathbf{o}_{\tilde{u}}, \sigma^*(u) = \tilde{u}\} = \frac{\mathbb{P}\Pr\{\mathbf{A}_u^t = r, \mathbf{o}_{\tilde{u}}^{1:t} \mid \sigma(u) = \tilde{u}\} \cdot \mathbb{P}\Pr\{\mathbf{o}_{\tilde{u}}^{t+1:T} \mid \mathbf{A}_u^t = r, \sigma(u) = \tilde{u}\}}{\mathbb{P}\Pr\{\mathbf{o}_{\tilde{u}} \mid \sigma^*(u) = \tilde{u}\}}$$

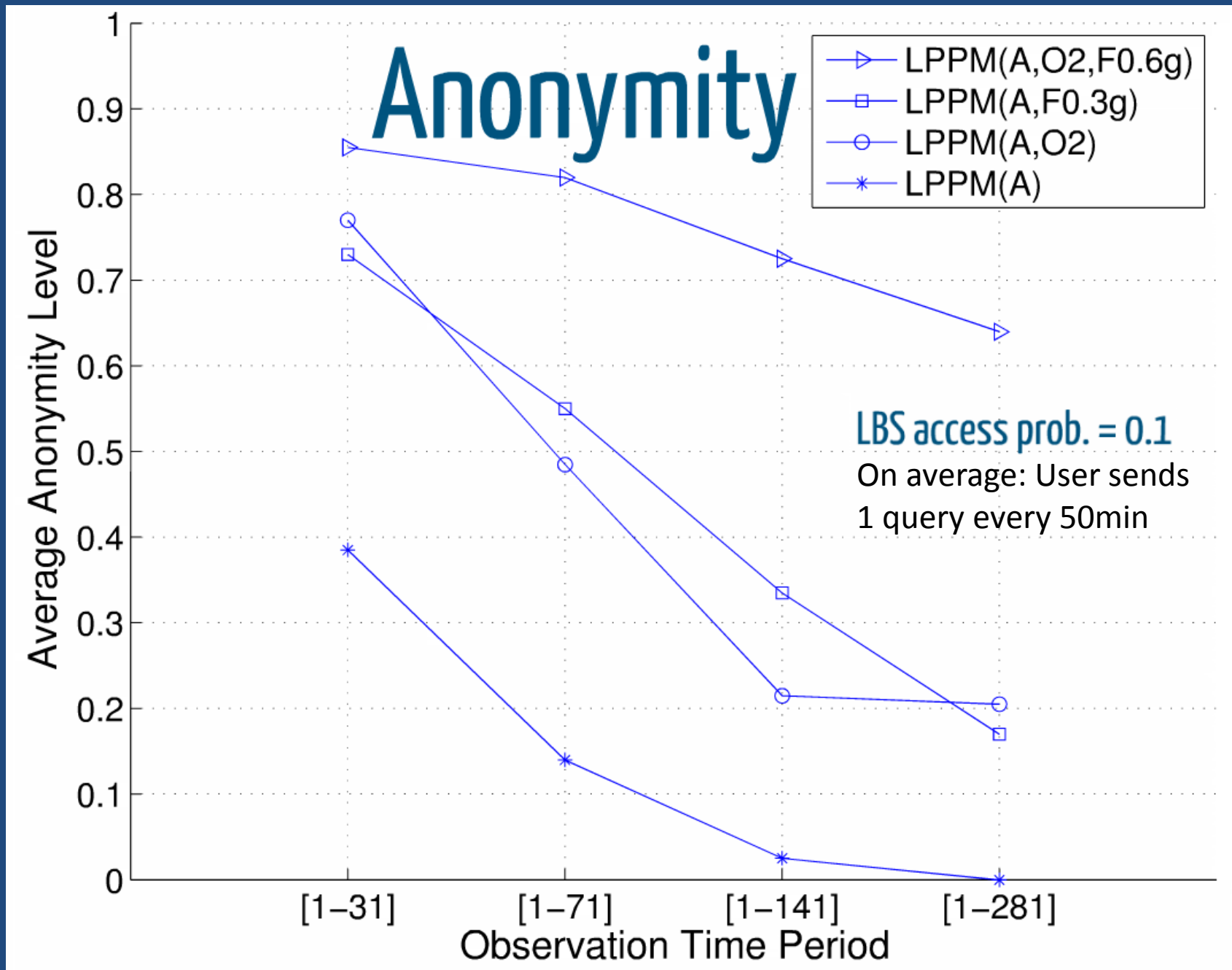
$$= \frac{\alpha_r^{u, \tilde{u}}(t) \cdot \beta_r^{u, \tilde{u}}(t)}{\sum_{r \in \mathcal{R}} \alpha_r^{u, \tilde{u}}(T)}$$



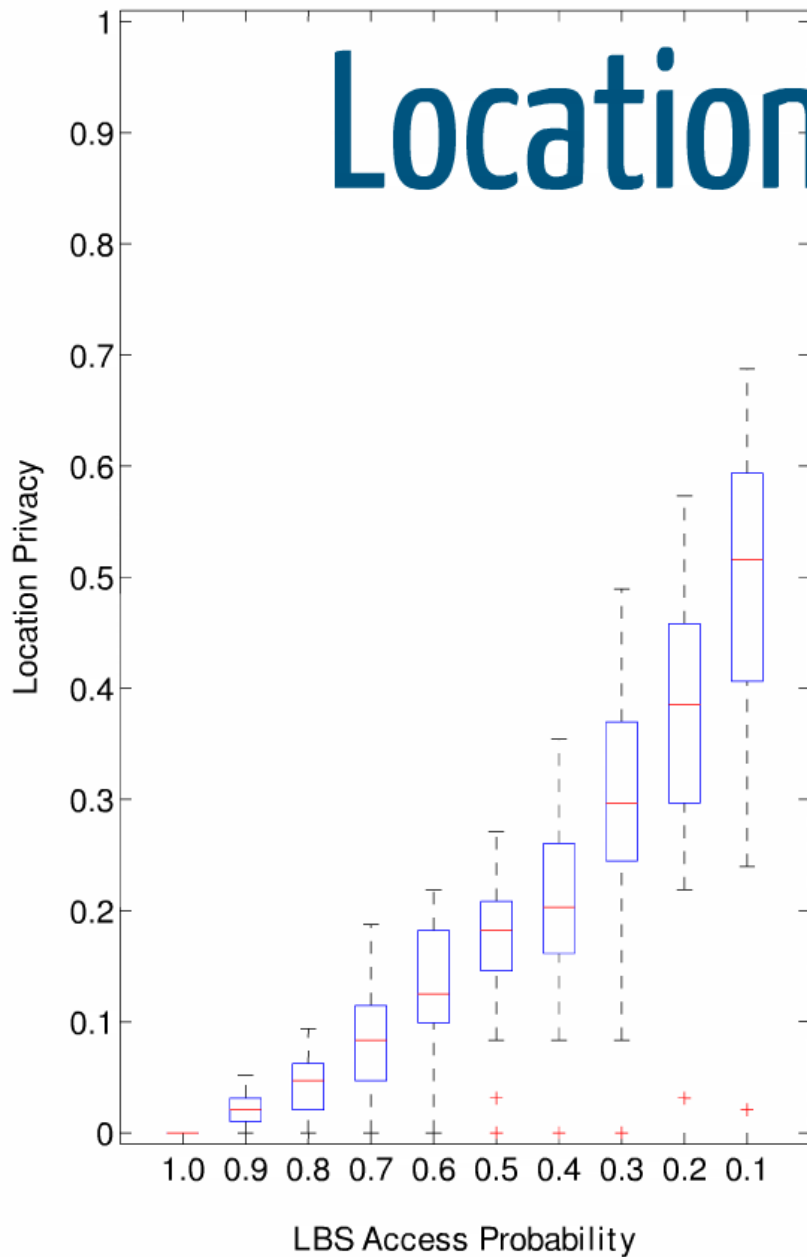
# Using LPM: Some Examples

- Real location traces
  - Time instant: 5min
  - 40 Locations in SF bay-area
- LBS Application
  - Sharing location with some access prob.  $p$  at each time instant
- LPPM
  - **A**: Anonymize (random permutation)
  - **On**: Obfuscate within  $2^n$  nearby locations
  - **Fm**: Send a fake location to LBS with prob.  $m$  (when user does not have a query herself)

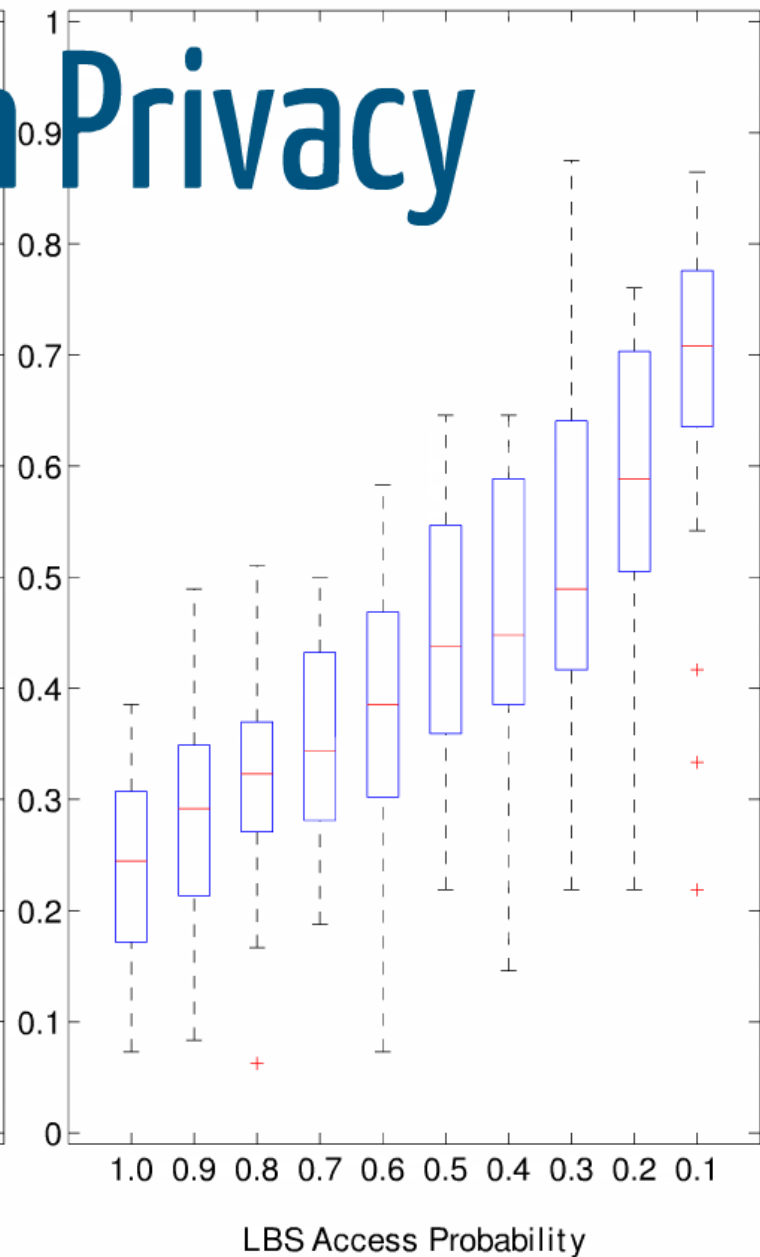




# Location Privacy



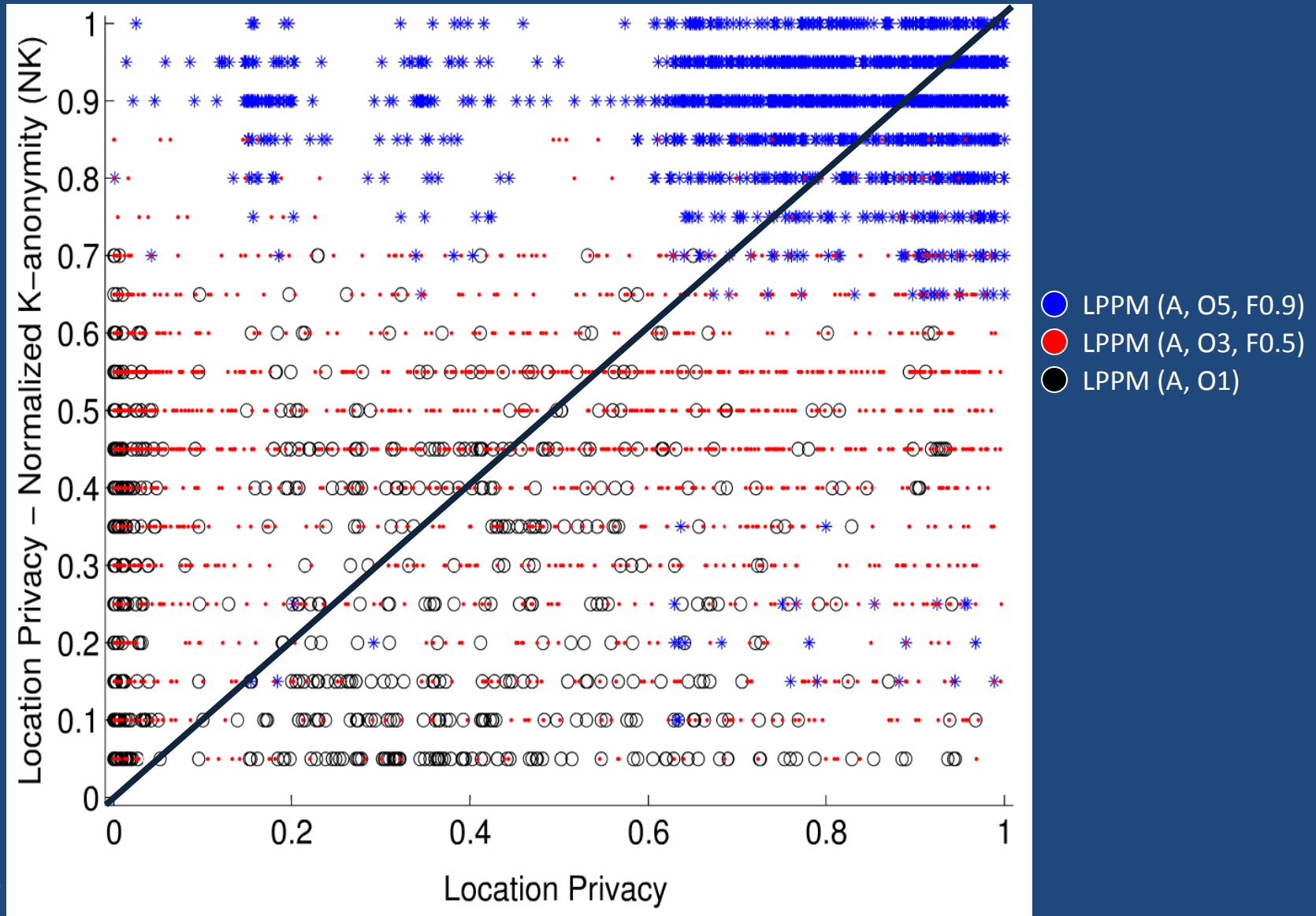
(a) LPPM(A)



(b) LPPM(A, O4)

Location Privacy: Adversary's probability of error in finding correct location, averaged over all locations.

# Evaluating Other Metrics: K-anonymity



# Conclusion and Remaining Issues

- We developed Location-Privacy Meter tool that enables us to consistently evaluate and compare **effectiveness** of location-privacy protection mechanisms (LPPMs), using Bayesian inference
- Yet, How to:
  - Maximize location privacy?
  - Find a balance between privacy and service quality?
  - Protect against a strategic adversary (best inference)?

# Protecting Location Privacy

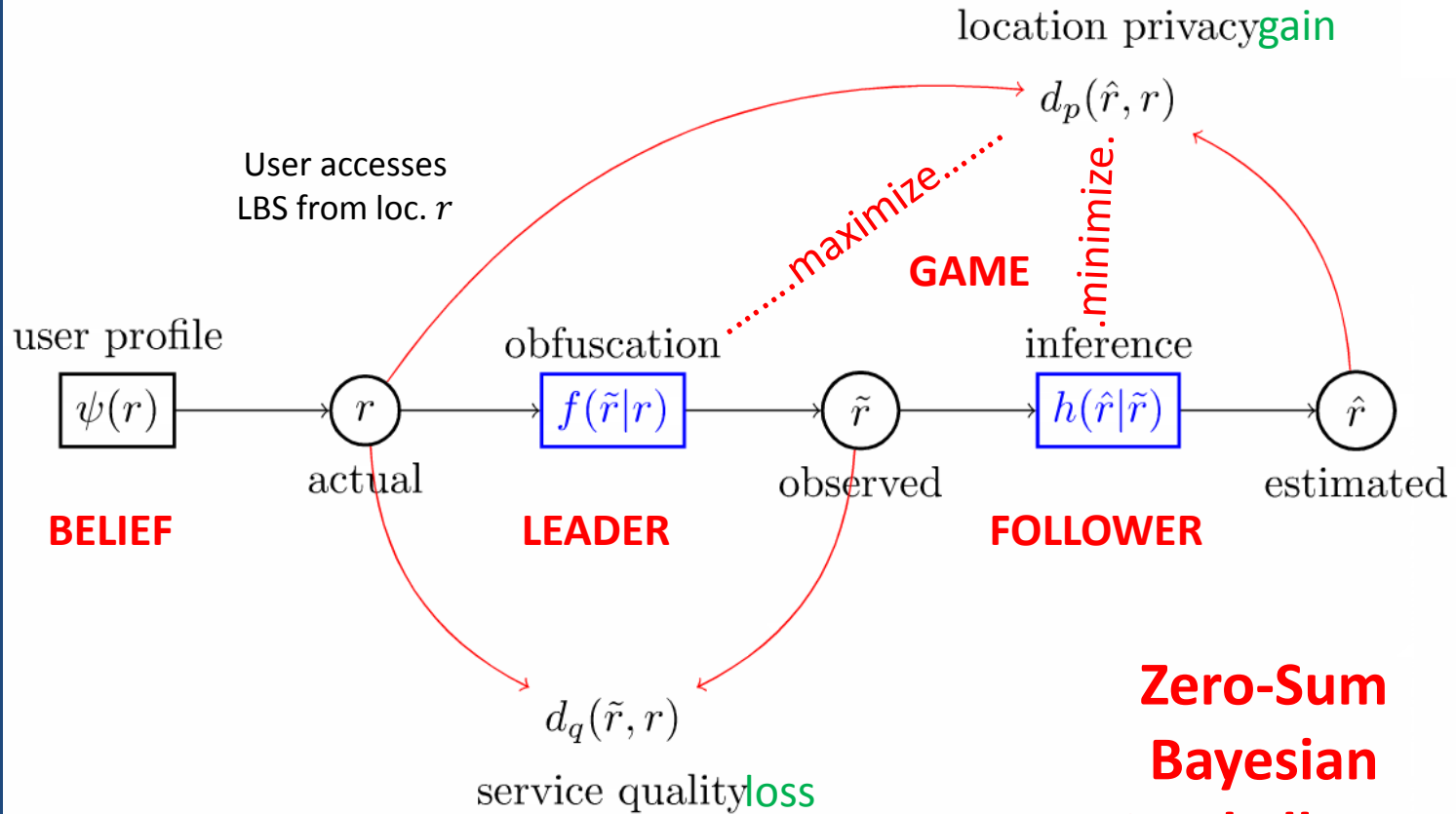
## Privacy Defense

# A User-Centric Approach

- Use our probabilistic model (introduced in the first part)
  - In modeling e.g., location, LBS, LPPM, and metric
- Respect each user's own privacy and service quality requirements
- Protect against the optimal inference attack, instead of assuming a given inference algorithm: Anticipate the location inference attacks
  - Each user protects against the strongest adversary that is specific to her (mobility and requirements)
- Model the Strategic interaction between user and attacker



# Game (Localization)



**Zero-Sum  
Bayesian  
Stackelberg  
Game**

$$Q_{loss}(\psi, f, d_q) = \sum_{r, \tilde{r}} \psi(r) \cdot f(\tilde{r}|r) \cdot d_q(\tilde{r}, r) \leq Q_{loss}^{\max}$$

$$Privacy(\psi, f, h, d_p) = \sum_{\hat{r}, \tilde{r}, r} \psi(r) \cdot f(\tilde{r}|r) \cdot h(\hat{r}|\tilde{r}) \cdot d_p(\hat{r}, r)$$

# Optimal Obfuscation

Choose  $f(\tilde{r}|r), x_{\tilde{r}}, \forall r, \tilde{r}$  in order to

Maximize  $\sum_{\tilde{r}} x_{\tilde{r}}$  **User's Privacy**  $\min_{\hat{r}} \sum_r \psi(r) \cdot f(\tilde{r}|r) \cdot d_p(\hat{r}, r)$

subject to

$$x_{\tilde{r}} \leq \sum_r \psi(r) \cdot f(\tilde{r}|r) \cdot d_p(\hat{r}, r), \forall \hat{r}, \tilde{r}$$

Respect user's  
service quality  
constraint

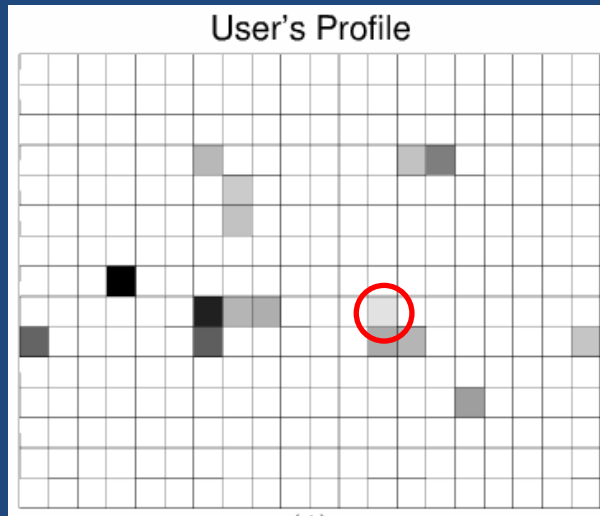
$$\sum_r \psi(r) \sum_{\tilde{r}} f(\tilde{r}|r) \cdot d_q(\tilde{r}, r) \leq Q_{loss}^{\max}$$

Proper probability  
distribution function

$$\sum_{\tilde{r}} f(\tilde{r}|r) = 1, \forall r$$

$$f(\tilde{r}|r) \geq 0, \forall r, \tilde{r}$$

# Visualization

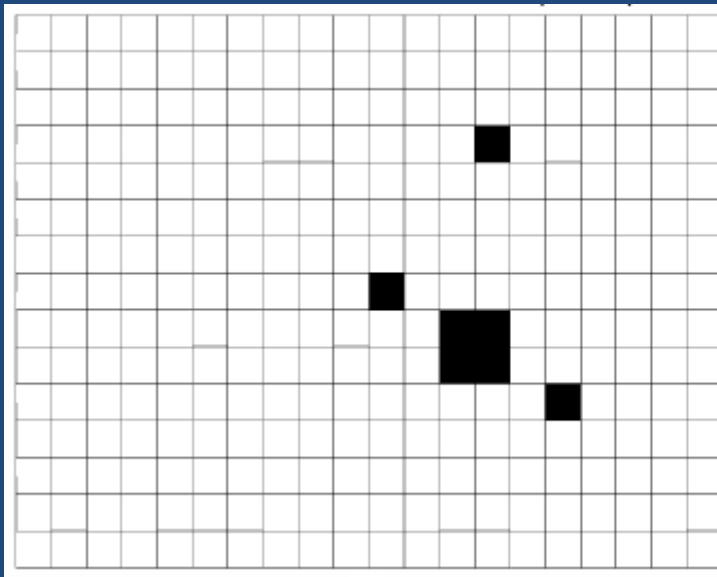


30 most visited locations

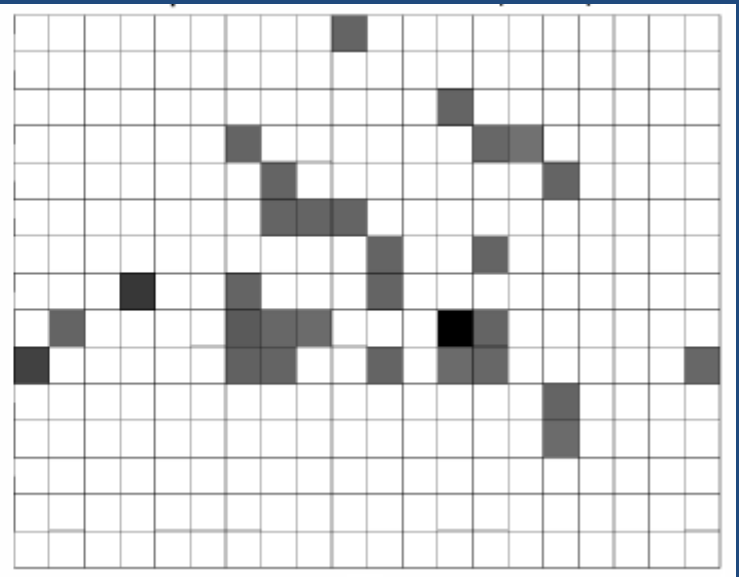
pdf: K-nearest Obfuscation

The service quality threshold of the optimal obfuscation function is set to the service quality loss of the k-nearest obfuscation function

pdf: Optimal Obfuscation



Uniform dist. over k nearest non-zero prob. neighbors



Distribution according to optimal LPPM f

# Optimal Inference

Dual of the optimal obfuscation LP

Choose  $h(\hat{r}|\tilde{r})$ ,  $y_r$ ,  $\forall r, \tilde{r}, \hat{r}$ , and  $z \in [0, \infty)$  in order to

**Minimize** 
$$\sum_r \psi(r) \cdot y_r + z \cdot Q_{loss}^{\max}$$

Minimizing the user's maximum privacy under the service quality constraint

**subject to**

$$y_r \geq \sum_{\hat{r}} h(\hat{r}|\tilde{r}) \cdot d_p(\hat{r}, r) - z \cdot d_q(\tilde{r}, r), \forall r, \tilde{r}$$

$$\sum_{\hat{r}} h(\hat{r}|\tilde{r}) = 1, \forall \tilde{r}$$

Proper probability distribution function

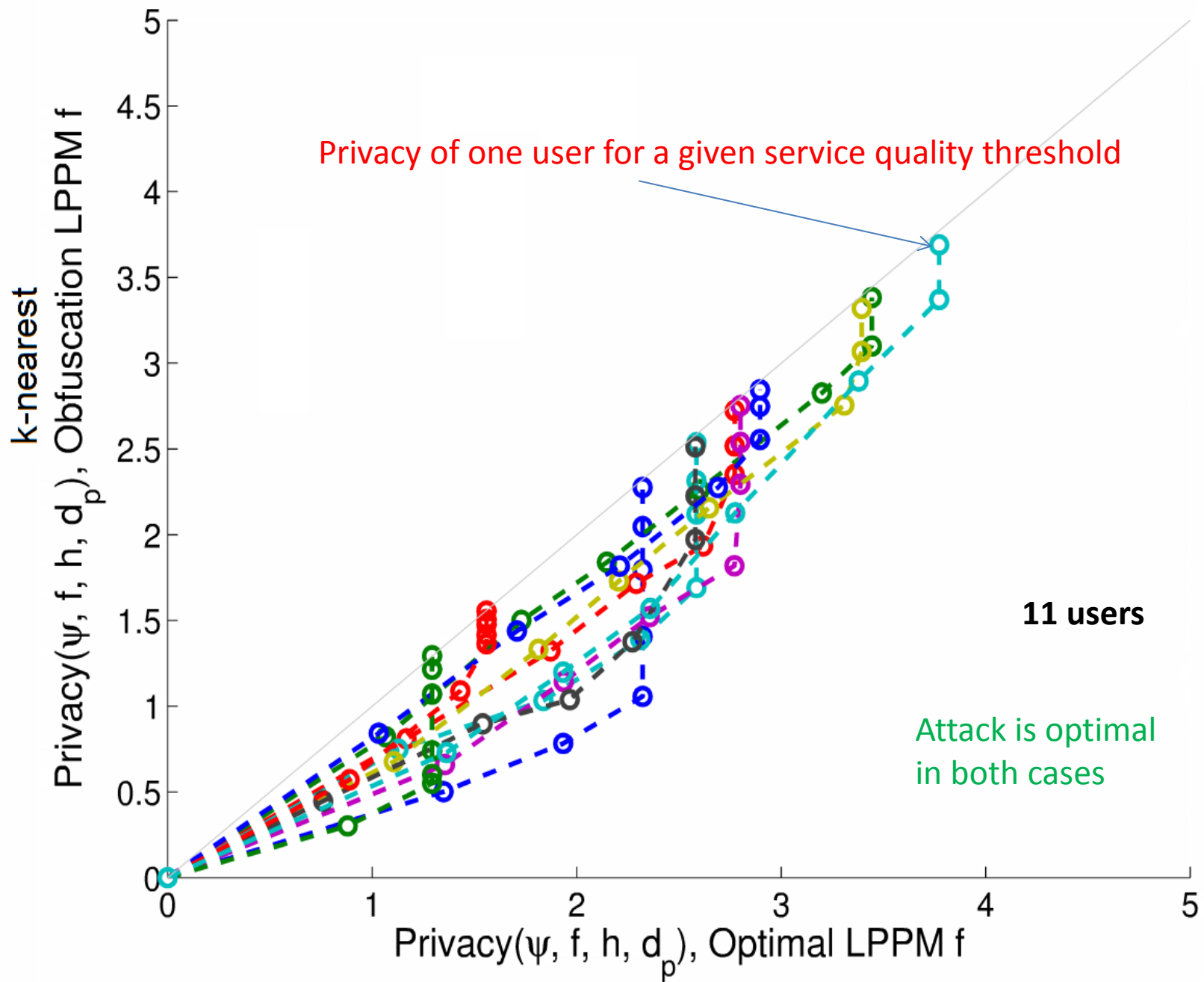
$$h(\hat{r}|\tilde{r}) \geq 0, \forall \tilde{r}, \hat{r}$$

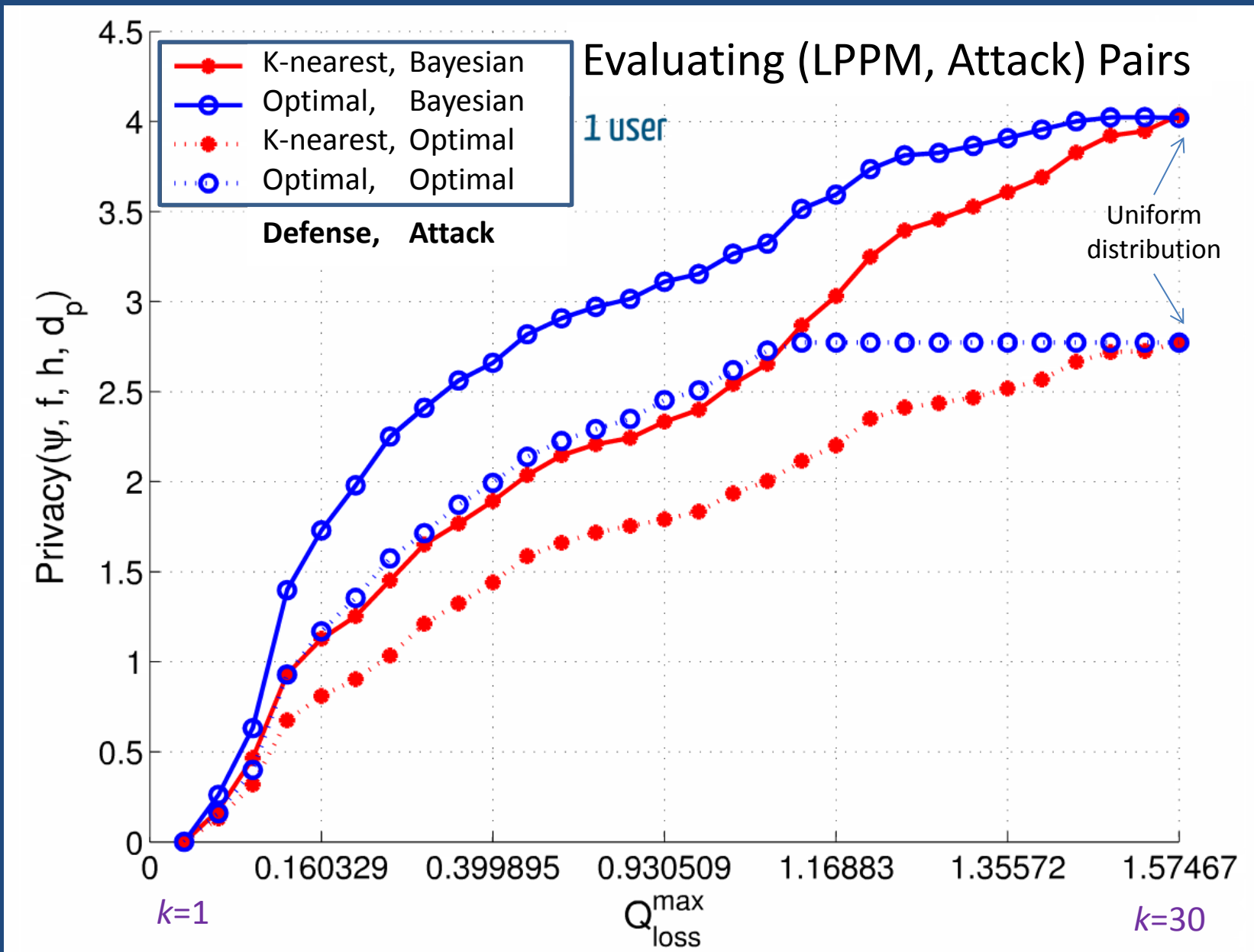
$$z \geq 0 \quad \left| \begin{array}{l} \text{Shadow price of the service quality constraint.} \\ \text{(exchange rate between service quality and privacy)} \end{array} \right.$$

# Evaluation:

## Optimal vs. Existing Methods

- Real location traces
  - Collected by **Nokia** Lausanne
- Obfuscation
  - K-nearest
  - Optimal
- Attack
  - Bayesian (not considering user's service quality constraints)
  - Optimal
- Metric
  - Both  $d_p$  and  $d_q$  are Euclidean distance functions





The Bayesian Inference Attack ignores the service quality constraint

# Conclusion

- We proposed an interactive decision making (game-theoretic) approach for protecting privacy in data-sharing applications
  - Anticipate inference attacks (rational adversary)
  - Respect user's service quality constraint
- Privacy risk is user-specific, hence should be the protection mechanisms



# Conclusion

- We need accurate models plus useful tools
- Users themselves are unable to accurately evaluate their privacy level and to define effective defenses
  - We provide tools to quantify and protect location privacy
- Privacy is user-dependent
  - Intelligent tools need to adapt to user's requirements
  - A user's behavior can be analyzed to learn her data model, sensitivities, and requirements (e.g., which places she visits, which checked-in locations she deletes later)
- Our Bayesian inference and game-theoretic approaches can be used in other data-sharing systems

# Acknowledgments

Vincent Bindschaedler, George Danezis, Claudia Diaz, Julien Freudiger, Jean-Pierre Hubaux, Mathias Humbert, Murtuza Jadliwala, Jean-Yves Le Boudec, Panos Papadimitratos, Pedram Pedarsani, Marcin Poturalski, Gael Ravot, Maxim Raya, Francisco Santos, George Theodorakopoulos, Carmela Troncoso