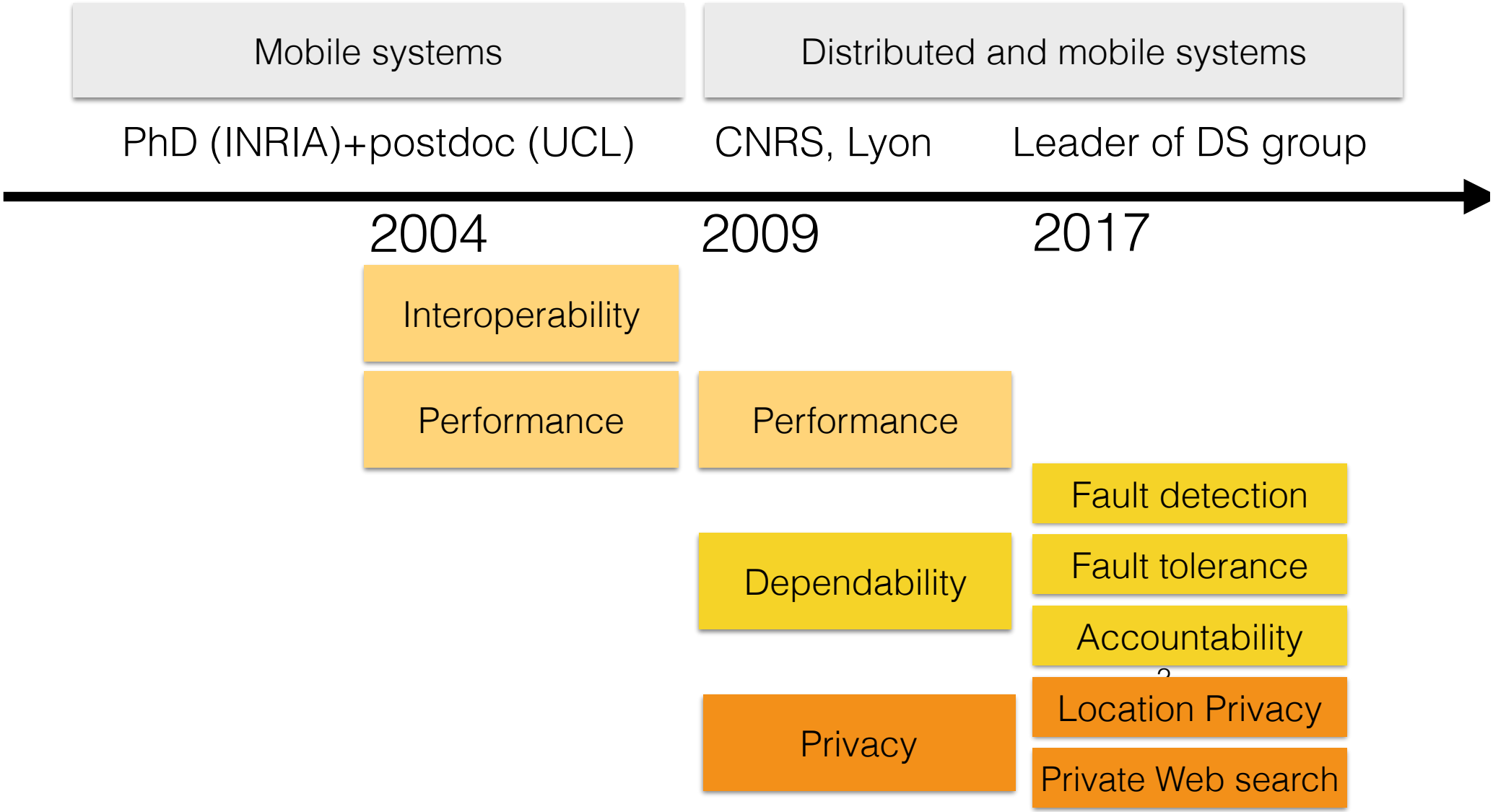
# Enforcing Privacy in Online Services

the privacy vs utility tradeoff

Sonia Ben Mokhtar

73rd IFIP Meeting
Goa, India
11th Jan. 2018

# Who am I?


Mobile systems
Distributed and mobile systems
PhD (INRIA)+postdoc (UCL)
CNRS, Lyon
Leader of DS group
2004
2009
2017
Interoperability
Performance
Performance
Fault detection
Dependability
Fault tolerance
Accountability
Privacy
Location Privacy
Private Web search

# Web Search

Every day, millions of users are querying **SEARCH ENGINES**

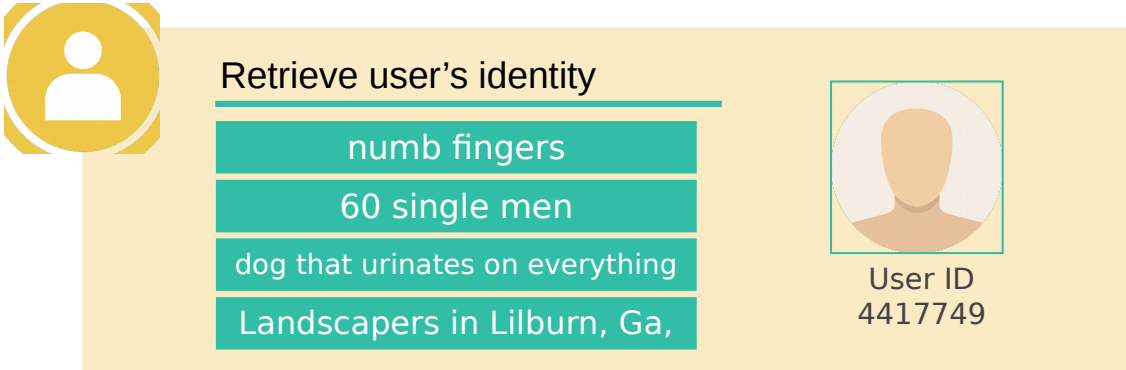
> We also use this information [*that we collect from all of our services*] to offer you tailored content – like giving you more **relevant search results** and **ads**.

*http://www.google.com/policies/privacy/*

YAHOO! bing g

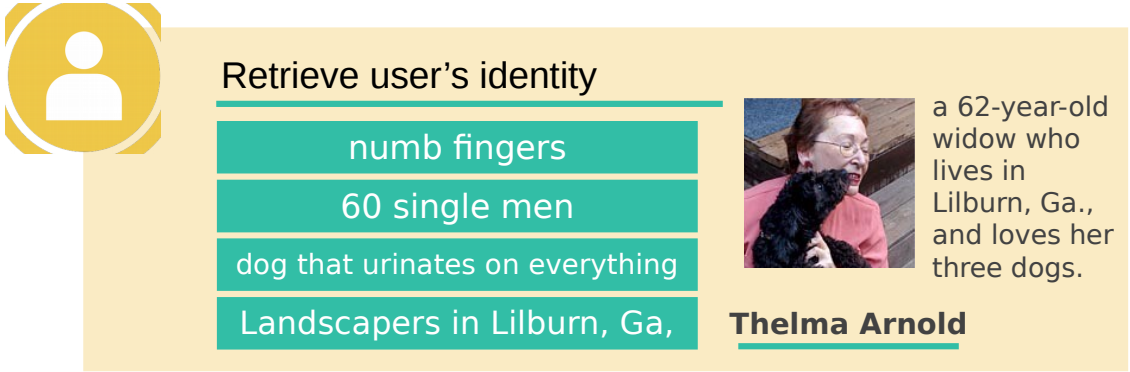
**USER PROFILES**

# Web Search: Privacy Threats




Barbaro, Michael, Tom Zeller, and Saul Hansell. "A face is exposed for AOL searcher no. 4417749." New York Times 9.2008 (2006): 8For.

# Web Search: Privacy Threats




Barbaro, Michael, Tom Zeller, and Saul Hansell. "A face is exposed for AOL searcher no. 4417749." New York Times 9.2008 (2006): 8For.

# Web Search: Privacy Threats

## Retrieve user's identity

- numb fingers
- 60 single men
- dog that urinates on everything
- Landscapers in Lilburn, Ga,

a 62-year-old widow who lives in Lilburn, Ga., and loves her three dogs.

**Thelma Arnold**

Barbaro, Michael, Tom Zeller, and Saul Hansell. "A face is exposed for AOL searcher no. 4417749." New York Times 9.2008 (2006): 8For.

## Infer extra information

- Age
- Zip Code
- Religion
- Gender
- Diseases
- Interests

Jones, Rosie, et al. "I know what you did last summer: query logs and user privacy." Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. ACM, 2007.

# Location-based Services

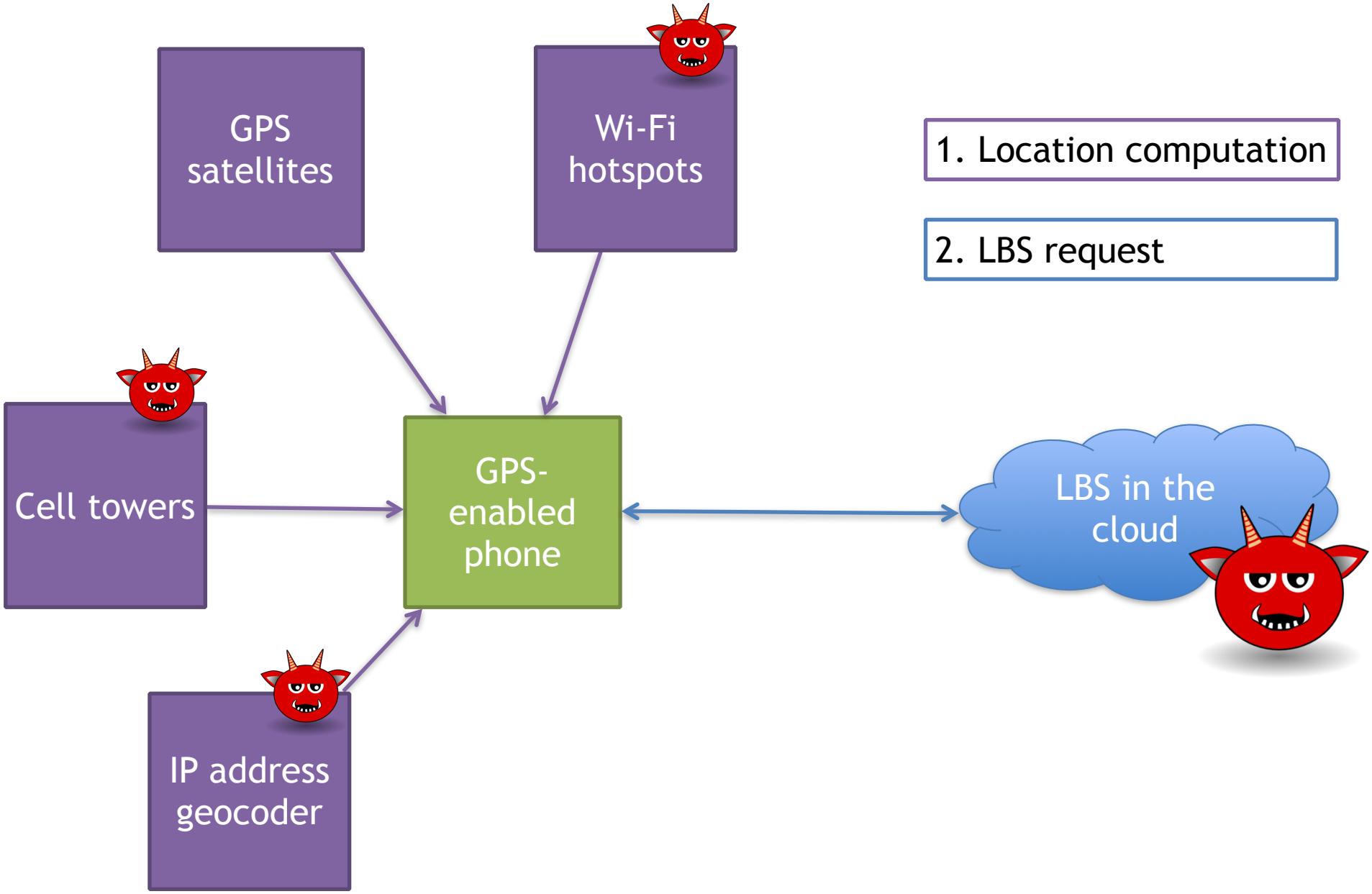
foursquare

Google maps


bing™ maps

OpenStreetMap

# Location lifecycle

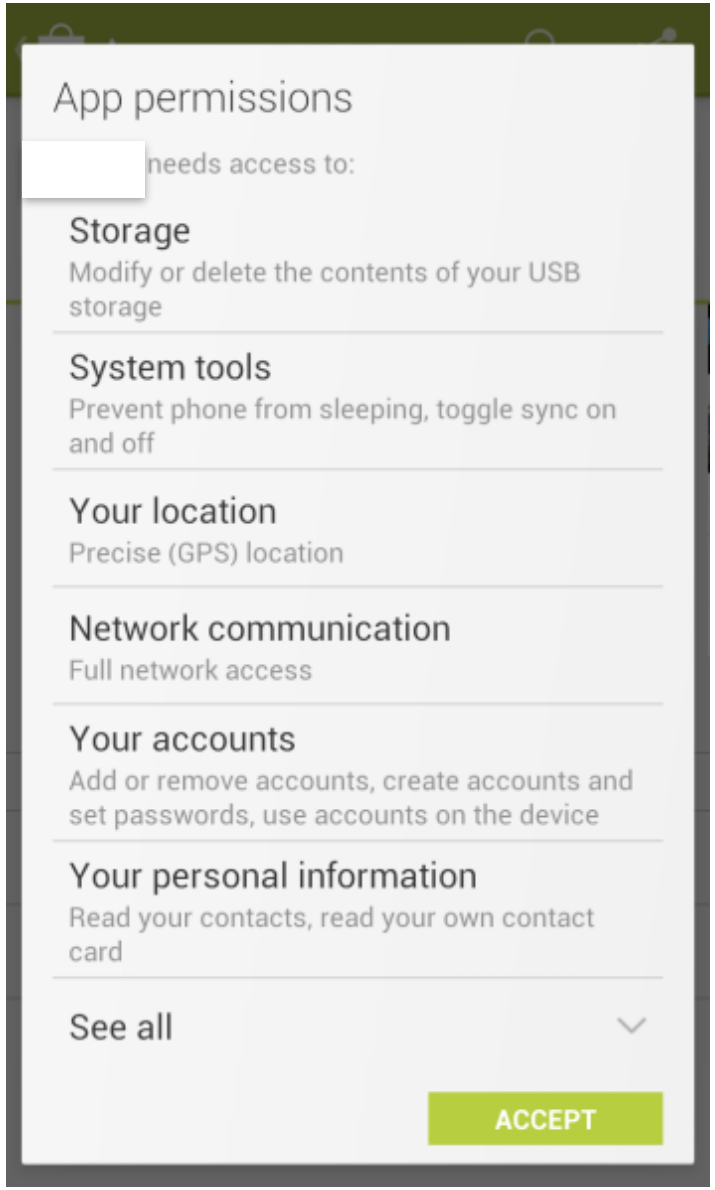


# Some numbers…

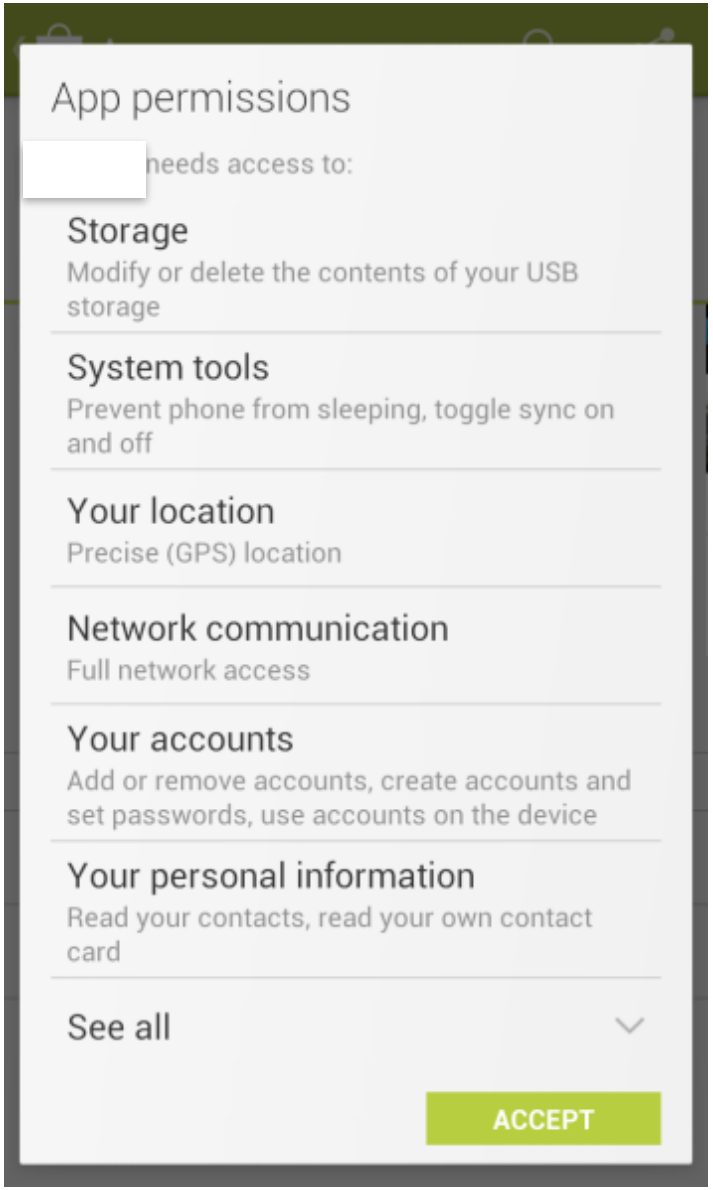
- Companies (e.g., Apple, TomTom…) have agreements to **share** location data with « partners and licensees »
- Skyhook wireless is **resolving** 400M user's WiFi locations/day
- <u>8B copies of applications downloaded from the AppStore **access location data**</u>
- ~50% of all iOS and Android traffic is available to ad networks

De Montjoye, Y.-A., Hidalgo, C., Verleysen, M. and Blondel, V. Unique in the Crowd: The privacy bounds of human mobility. Scientific reports,Scientific Reports 3, Article number: 1376, 2013.

# In practice…


App permissions
needs access to:
Storage
Modify or delete the contents of your USB storage
System tools
Prevent phone from sleeping, toggle sync on and off
Your location
Precise (GPS) location
Network communication
Full network access
Your accounts
Add or remove accounts, create accounts and set passwords, use accounts on the device
Your personal information
Read your contacts, read your own contact card
See all
ACCEPT

# In practice…

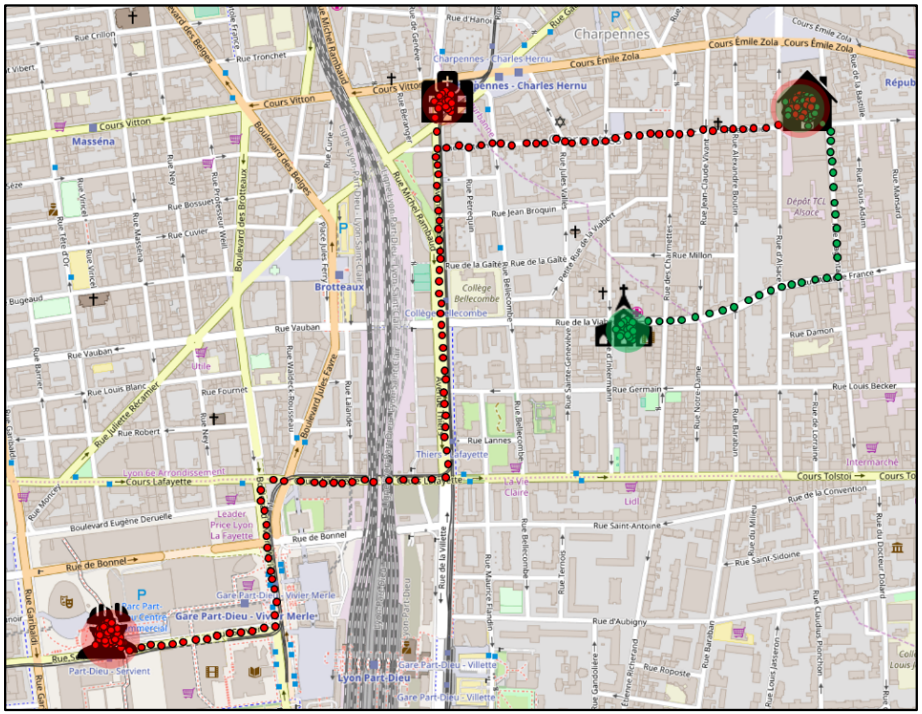

App permissions
needs access to:
Storage
Modify or delete the contents of your USB storage
System tools
Prevent phone from sleeping, toggle sync on and off
Your location
Precise (GPS) location
Network communication
Full network access
Your accounts
Add or remove accounts, create accounts and set passwords, use accounts on the device
Your personal information
Read your contacts, read your own contact card
See all
ACCEPT


LED LENSER
P14

# LBS: Privacy Threats

**From user mobility, sensitive information can be inferred :**

- **Points of Interest (POIs). Such as Home location, Work place, Place of worship.**
- **Social relationships. Such as Siblings, Coworkers, Partners.**
- **Re-identification.**
- **Mobility Prediction.**



# Privacy threats caused by online services

**Dropbox**

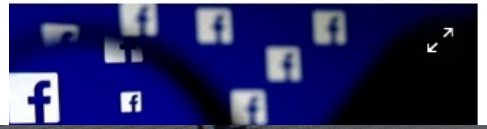
## Dropbox hack leads to leaking of 68m user passwords on the internet

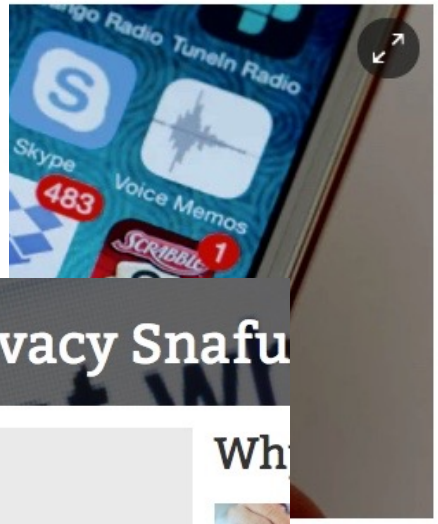
Data stolen in 2012 breach, containing encrypted passwords and details of ers, has been leaked

**Facebook**

## Revealed: Facebook exposed identities of moderators to suspected terrorists

d more than 1,000 workers forced one moderator into constant fear for his safety



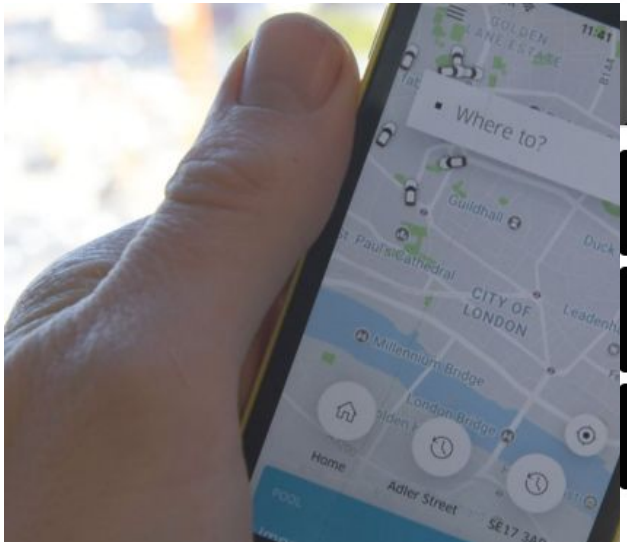



Technology

## Uber concealed huge data breach

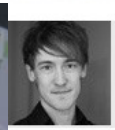
Dave Lee
North America technology reporter

22 November 2017 | Technology 173




24 APR 2017 NEWS

## LinkedIn Apologizes After Privacy Snafu



Phil Muncaster UK / EMEA News Reporter , Infosecurity Magazine
Email Phil Follow @philmuncaster

LinkedIn has apologized after its latest iOS update prompted some users to OK a new feature designed to connect them to nearby strangers within Bluetooth range.

The privacy snafu was spotted by Trend Micro VP of global research, Rik Ferguson, who claimed the update was described by LinkedIn merely as containing "general bug fixes and performance improvements."

Replying to his post on Twitter, several other users claimed to have been presented with the same pop-up following their download of the update.

It read: "LinkedIn would like to make data available to nearby Bluetooth devices even when you're not using the app. We will help you connect with others that are nearby."

The privacy implications of clicking "OK" are pretty obvious, and user anger at the appearance of the new feature seems to have been compounded by the complete lack of information

Wh

users' email

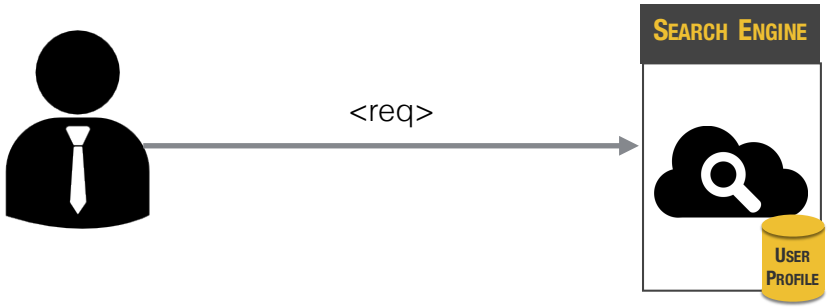
18 MA

How llection of
About rds had been

# European Law Enforcement

**The General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) requires Privacy by design. The Law enforces data protection and privacy of European citizens with severe penalties of up to 4% of the company's worldwide turnover.**

# Privacy in online services




- Classical scenario:
  - User sends request to the online service (e.g., search engine, location-based service, recommender system)
  - (Request may disclose sensitive information about the user)
  - Online service **gathers** this information (in the form of a user profile) and may **use it** (for improving the service) and/or leak it to third parties

# Existing solutions

- Generally solutions are devised for specific types of online services:
  - Private Web search
  - Location privacy
  - Private recommender systems
  - Private advertising
  - etc.

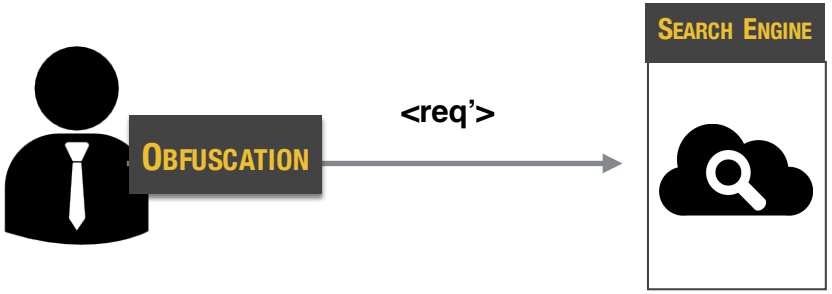
# Existing solutions

- Enforce known privacy properties
  - Differential privacy
  - K-anonymity and its variants
- Rely on practical privacy metrics
  - Resilience to re-identification attacks

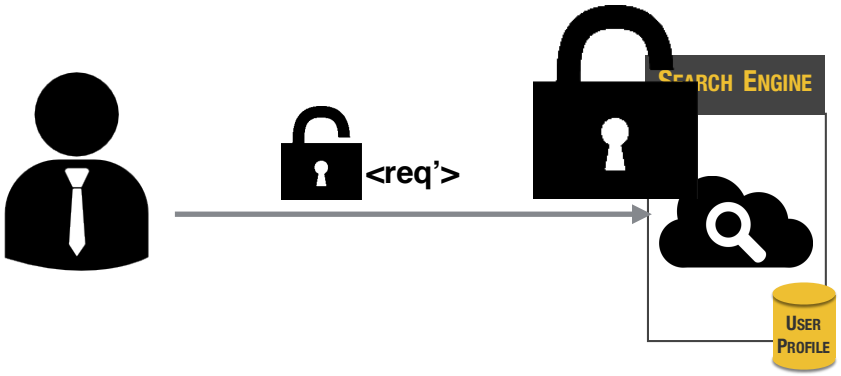
# Existing solutions

According to their architecture

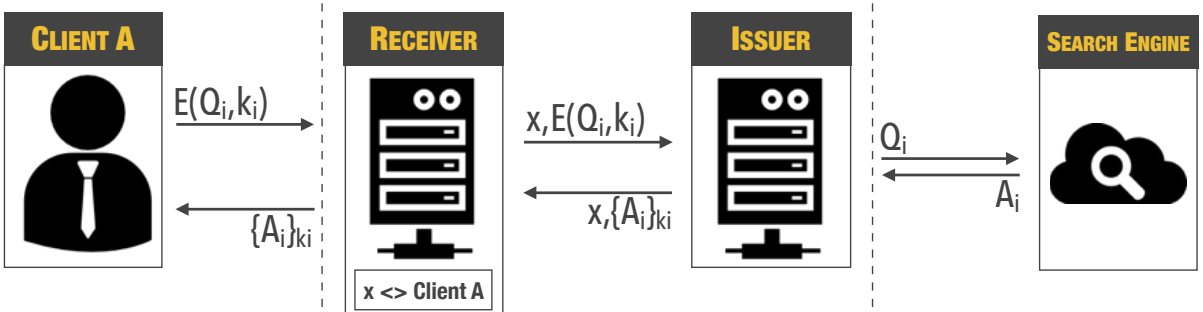
- **Client-side solutions**




- Server side solutions

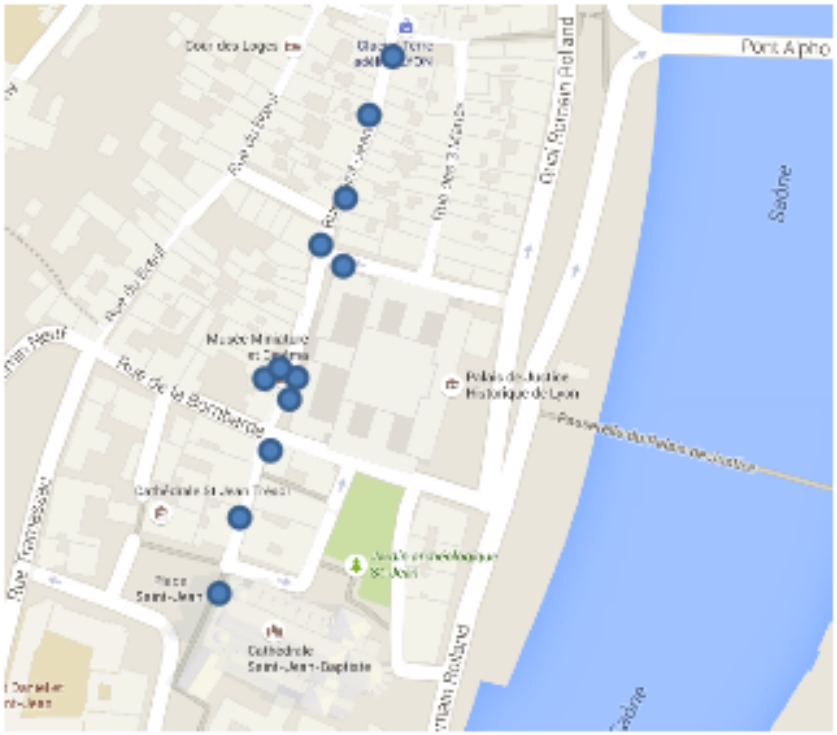



- **Proxy-based solutions**



# Location Privacy Protection Mechanisms -Promesse-

## Geo-indistinguishability




ε: amount of noise

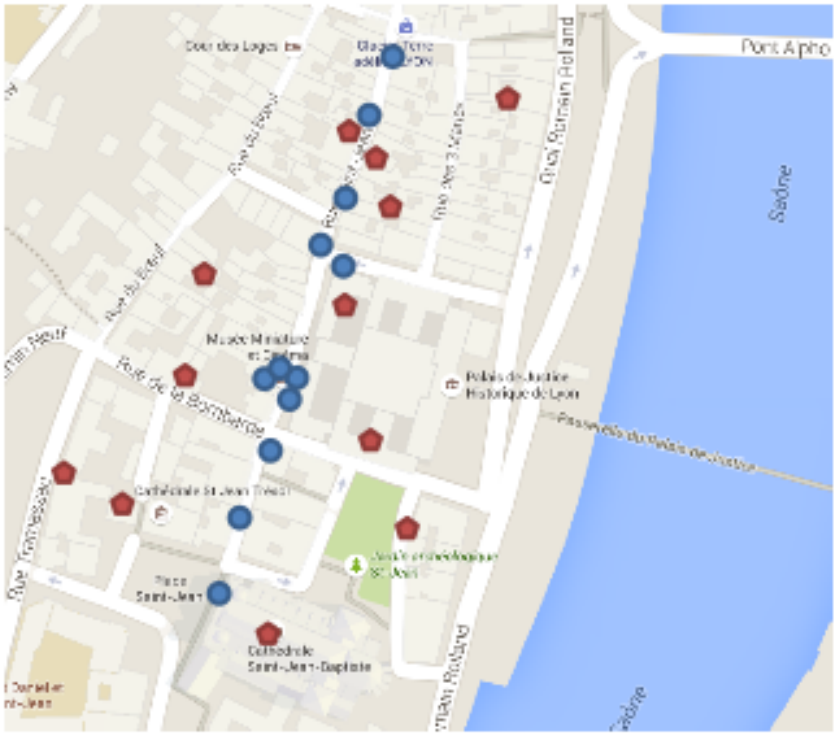
Actual location



Geo-indistinguishability: differential privacy for location-based systems. ACM CCS'13

# Location Privacy Protection Mechanisms -Promesse-

## Geo-indistinguishability

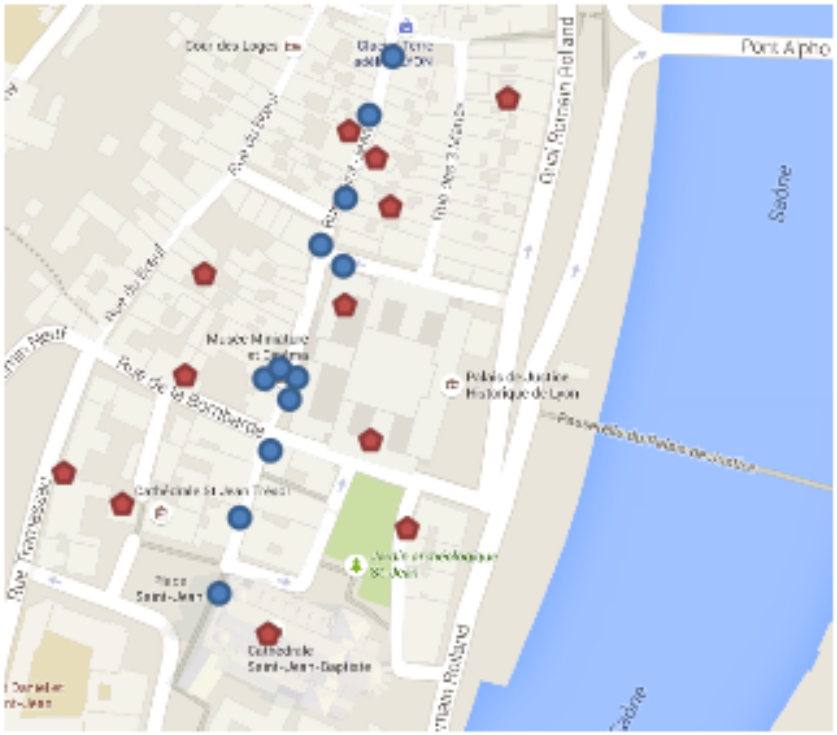


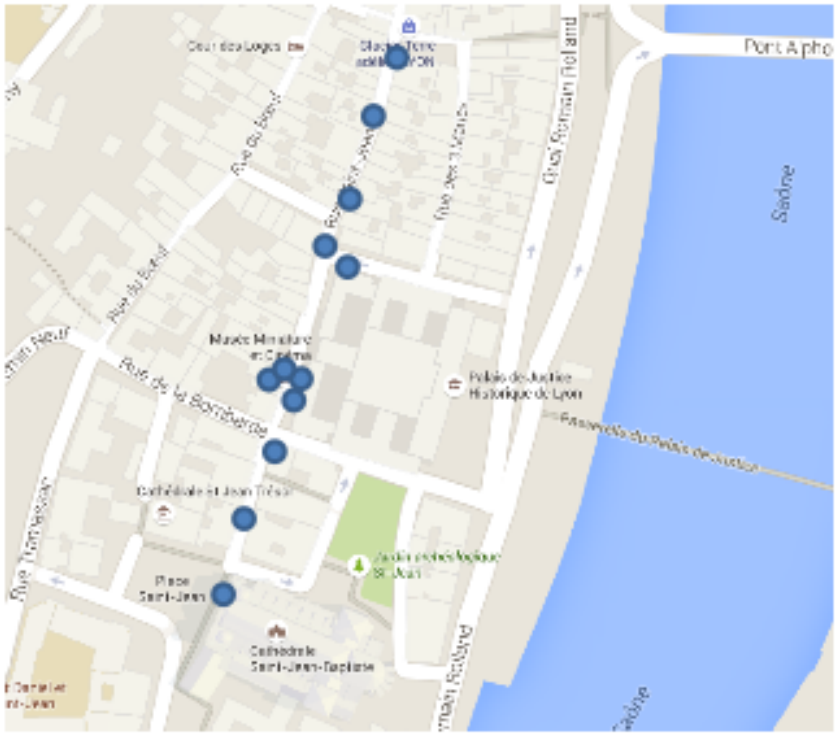

ε: amount of noise

Actual location  Protected location

# Location Privacy Protection Mechanisms -Promesse-

**Geo-indistinguishability** | **Promesse**



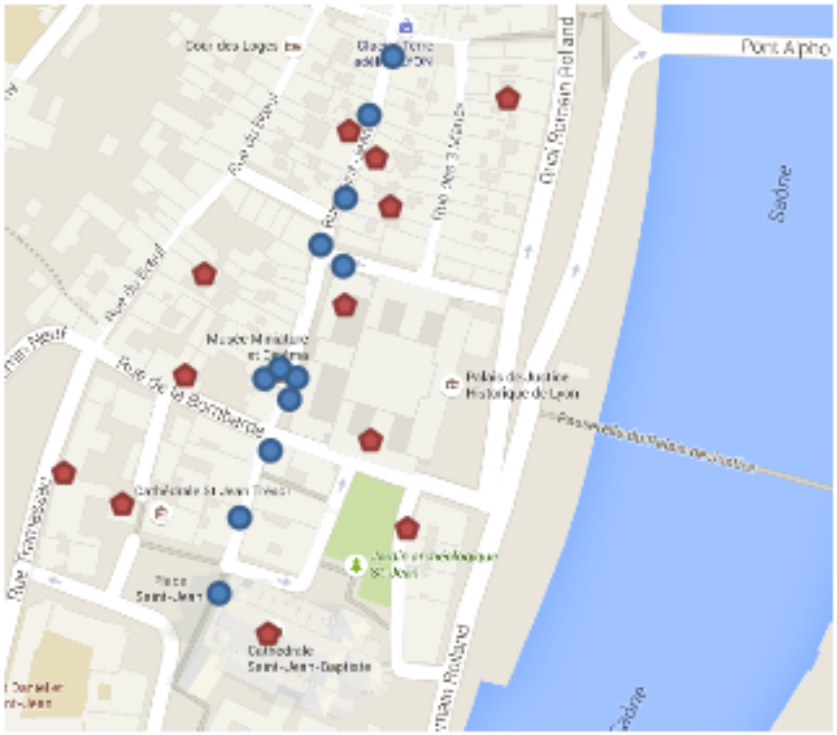



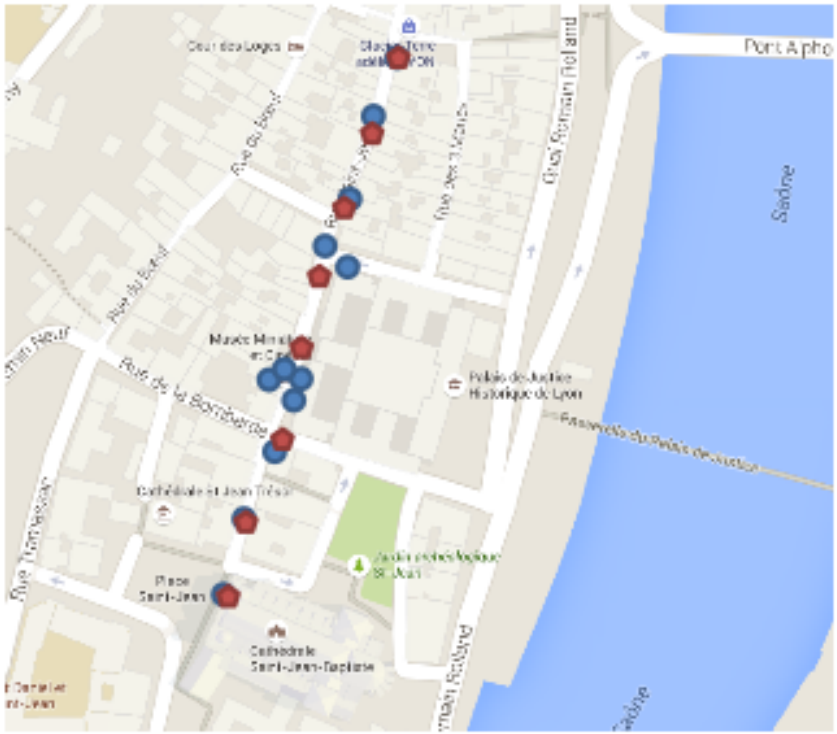

ε: amount of noise

Actual location  Protected location

# Location Privacy Protection Mechanisms -Promesse-

**Geo-indistinguishability** | **Promesse**





ε: amount of noise

α: distance between points

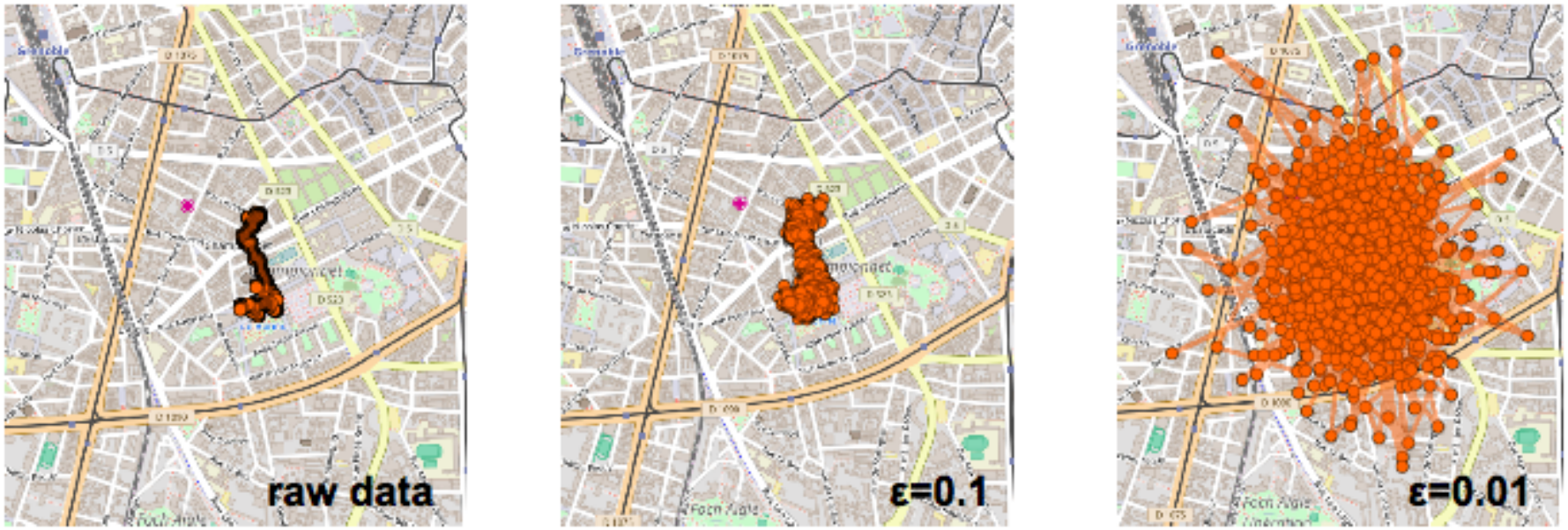
Actual location  Protected location



Time Distortion Anonymization for the Publication of Mobility Data with High Utility. IEEE TrustCom'15. 2015.

# Location Privacy Protection Mechanisms
# -Privacy vs Utility Trade-off-

- Geo Indistinguishability (**Geol**)
  - ε in meters$^{-1}$

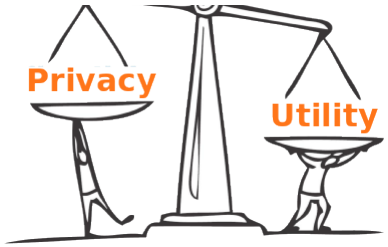


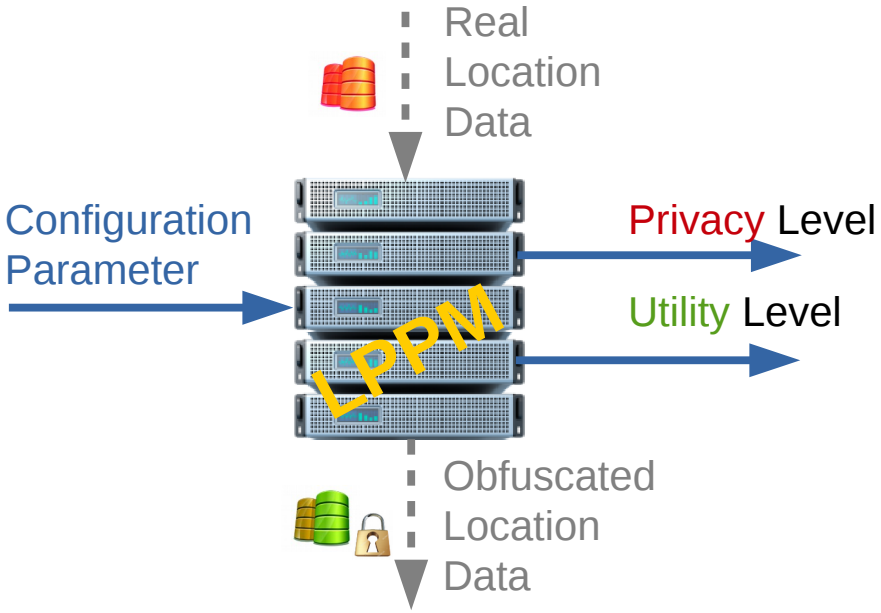



Geo-indistinguishability: differential privacy for location-based systems. ACM CCS'13

# LPPM configuration

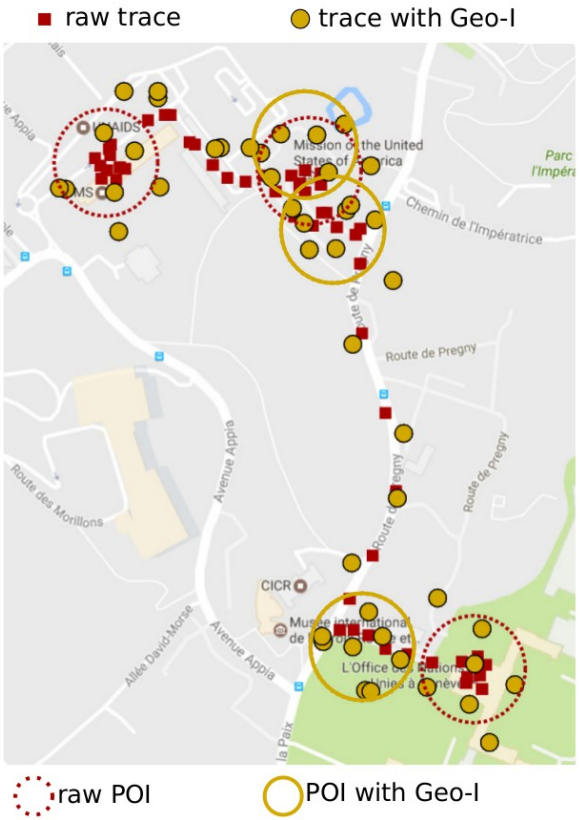
- How to measure privacy ? utility ?
- What is the **trade-off** between utility and privacy ?
- How to get privacy and utility **guarantees ?**



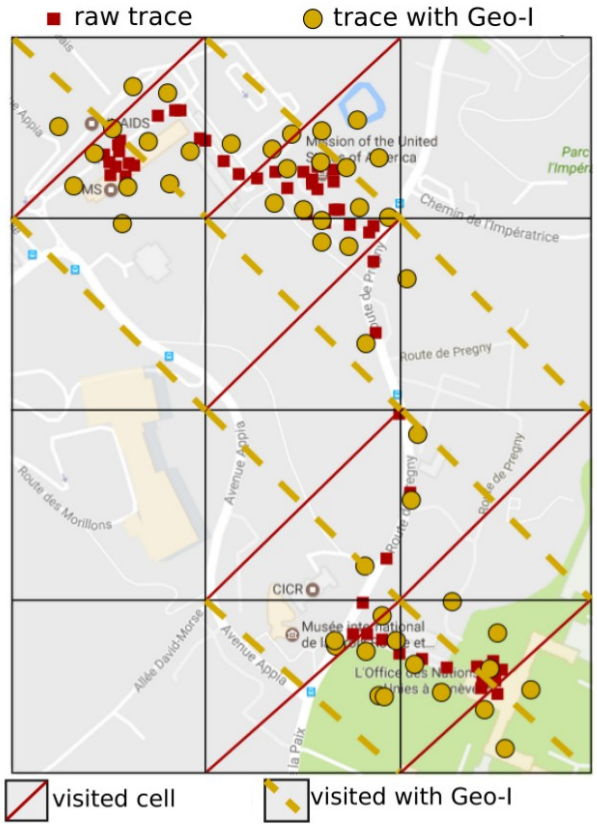



# Metrics

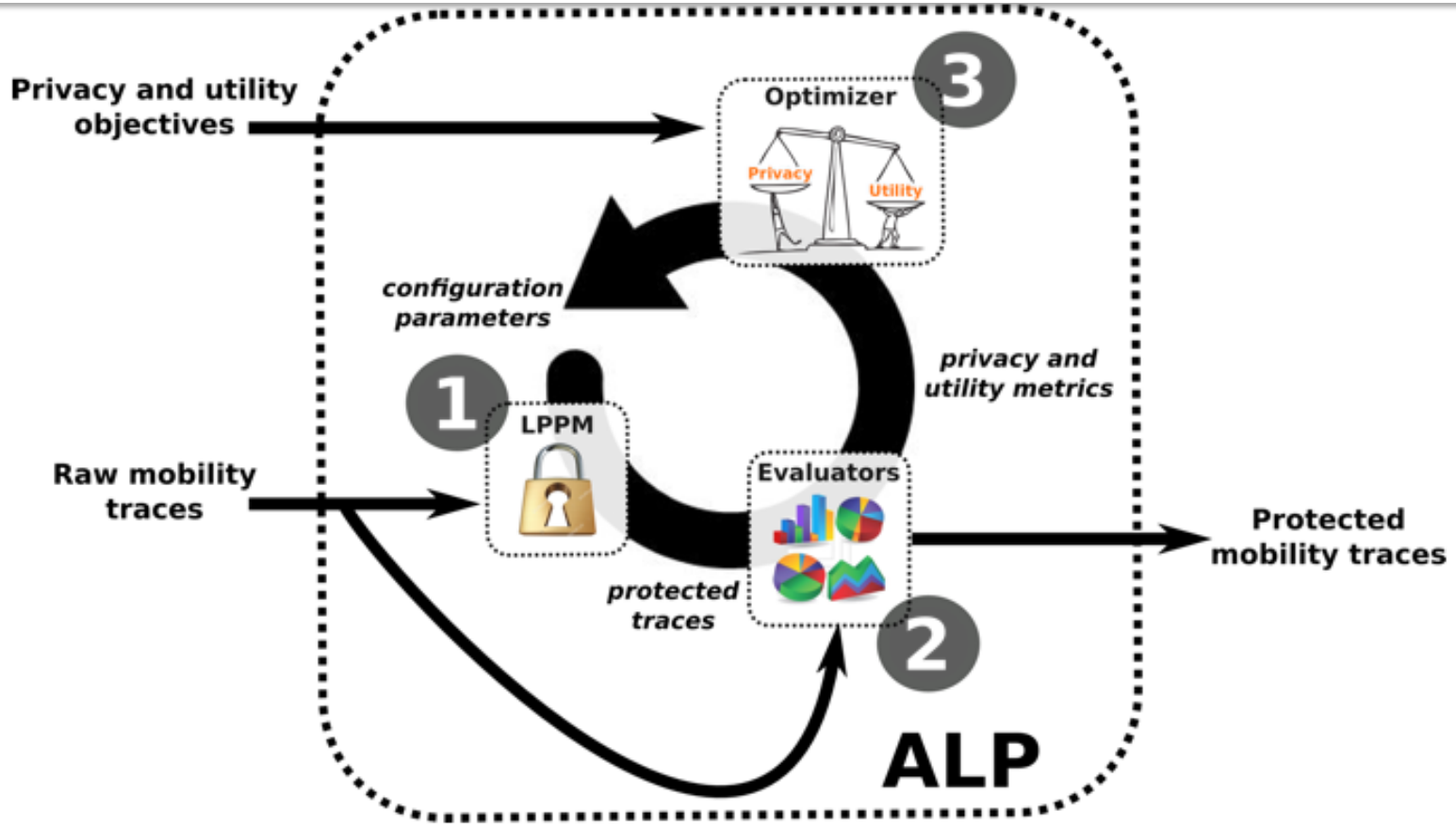
- Privacy ρ: proportion of hidden Points of Interest




- Utility μ: proportion of areas rightly covered



# Location Privacy Protection Mechanisms -ALP-



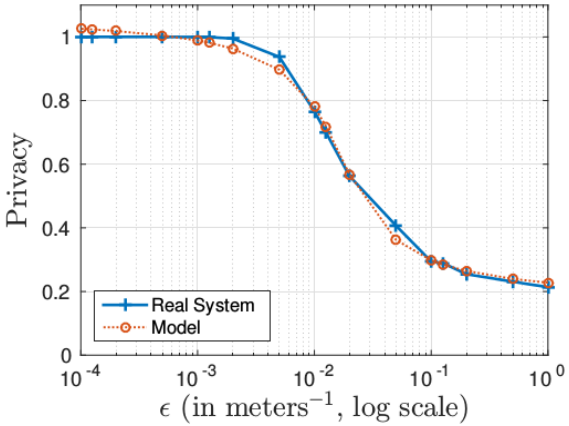


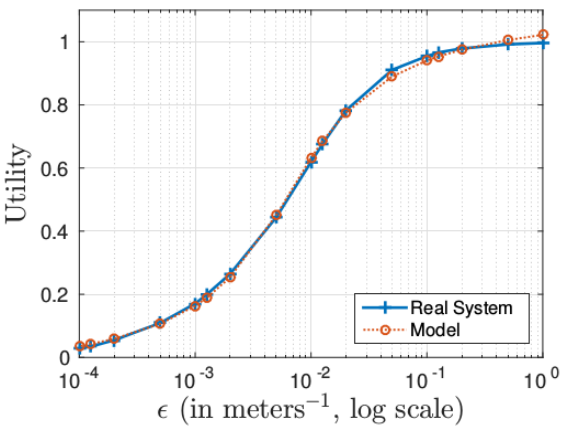



Adaptive Location Privacy with ALP. IEEE SRDS'16

# PULP: Modeling LPPMs

- **Objective**: Model the impact of an LPPM on a database




$$\rho = a_\rho.tan^{-1}\left(b_\rho(ln(\epsilon) - c_\rho)\right) + d_\rho$$

$$\mu = a_\mu.tan^{-1}\left(b_\mu(ln(\epsilon) - c_\mu)\right) + d_\mu$$

**Parameters adaptation to fit the dataset**

## EVALUATION

Modeling of 4 datasets: Mobility Data Challenge, Cabspotting, Geolife, Privamov

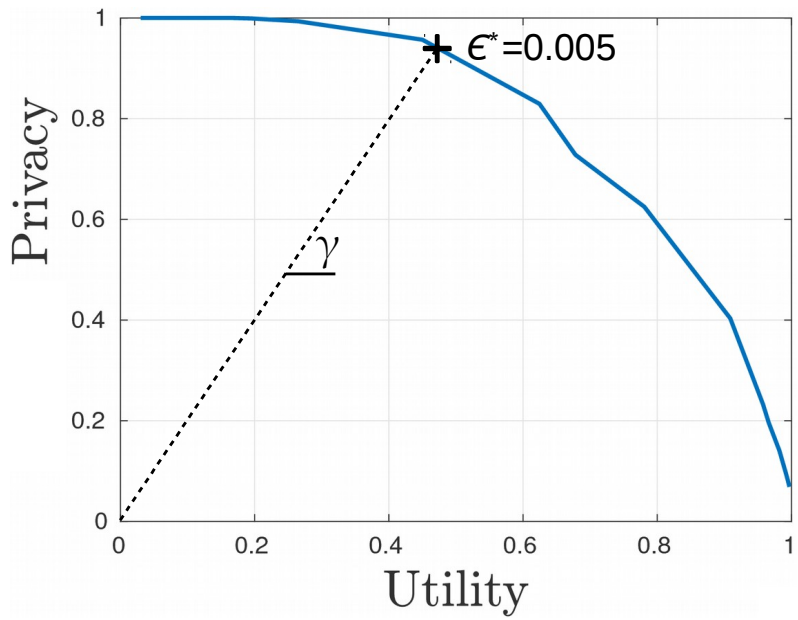
**Error Variance < $10^{-3}$**

# PULP: Configuring LPPMs

- **Objective**: Find adequate LPPM configuraton to achieve privacy to utility trade-off


$$\rho = \gamma\mu$$

$$\varepsilon^* = \arg\min_{\varepsilon} |\rho - \gamma\mu|$$

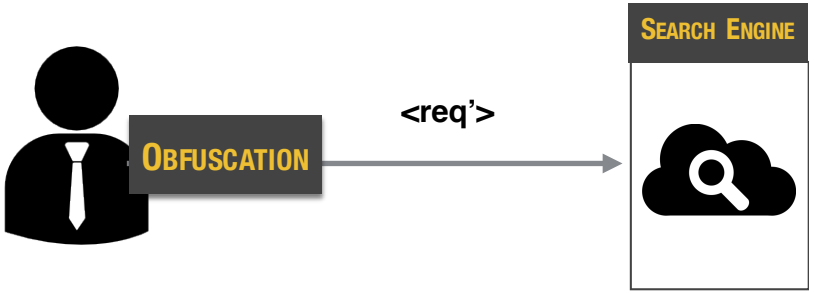


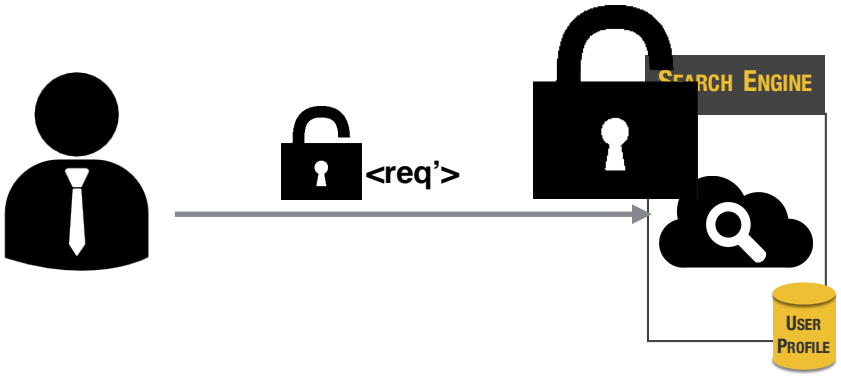

## EVALUATION

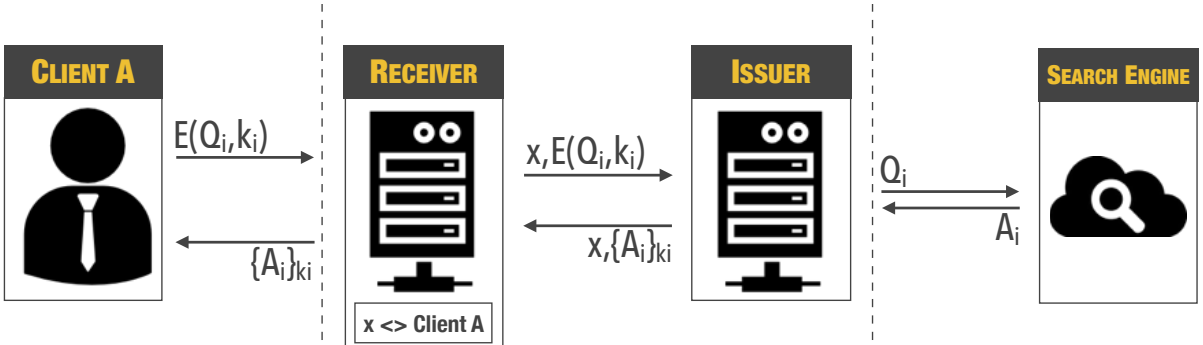
The error between desired trade-off $\gamma$ and the obtained one is bounded by the model error

# Existing solutions

- According to their architecture
  - Client-side solutions
  - Server side solutions
  - **<u>Proxy-based solutions</u>**



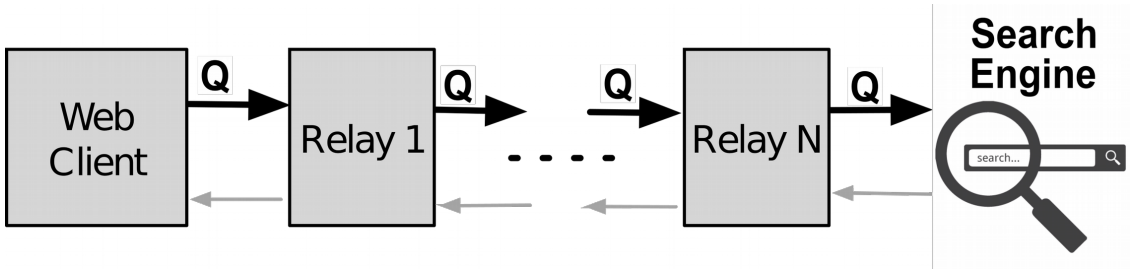
# Private Web search

- How can users protect their privacy from curious search engines?

1. Hiding identities (IP Address)
2. Making queries and user’s interests indistinguishable

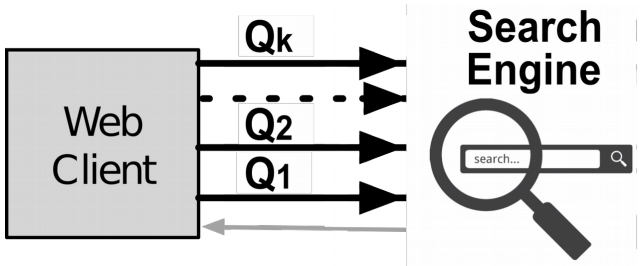
# Unlinkability

**Unlinkability** between user and query (Tor)

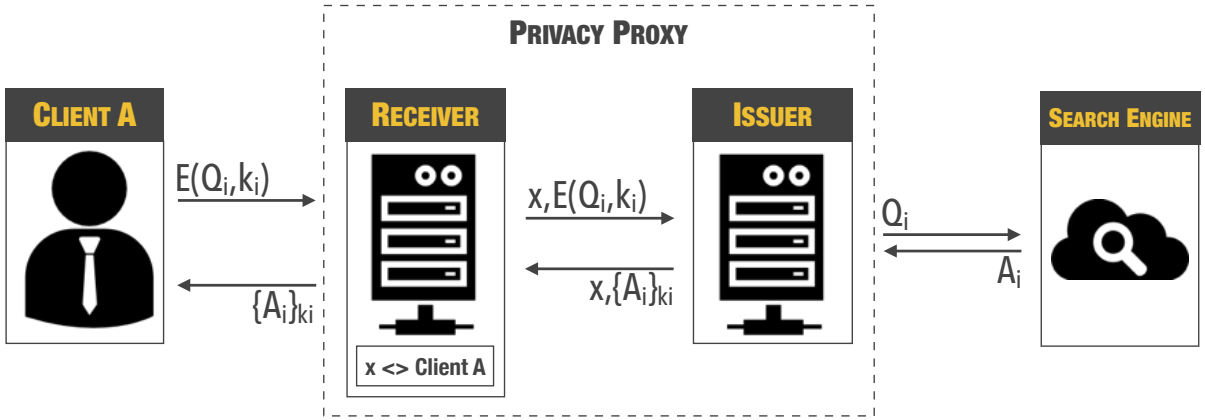


# Indistinguishability

**Indistinguishability** between real and fake queries (TrackMeNot)



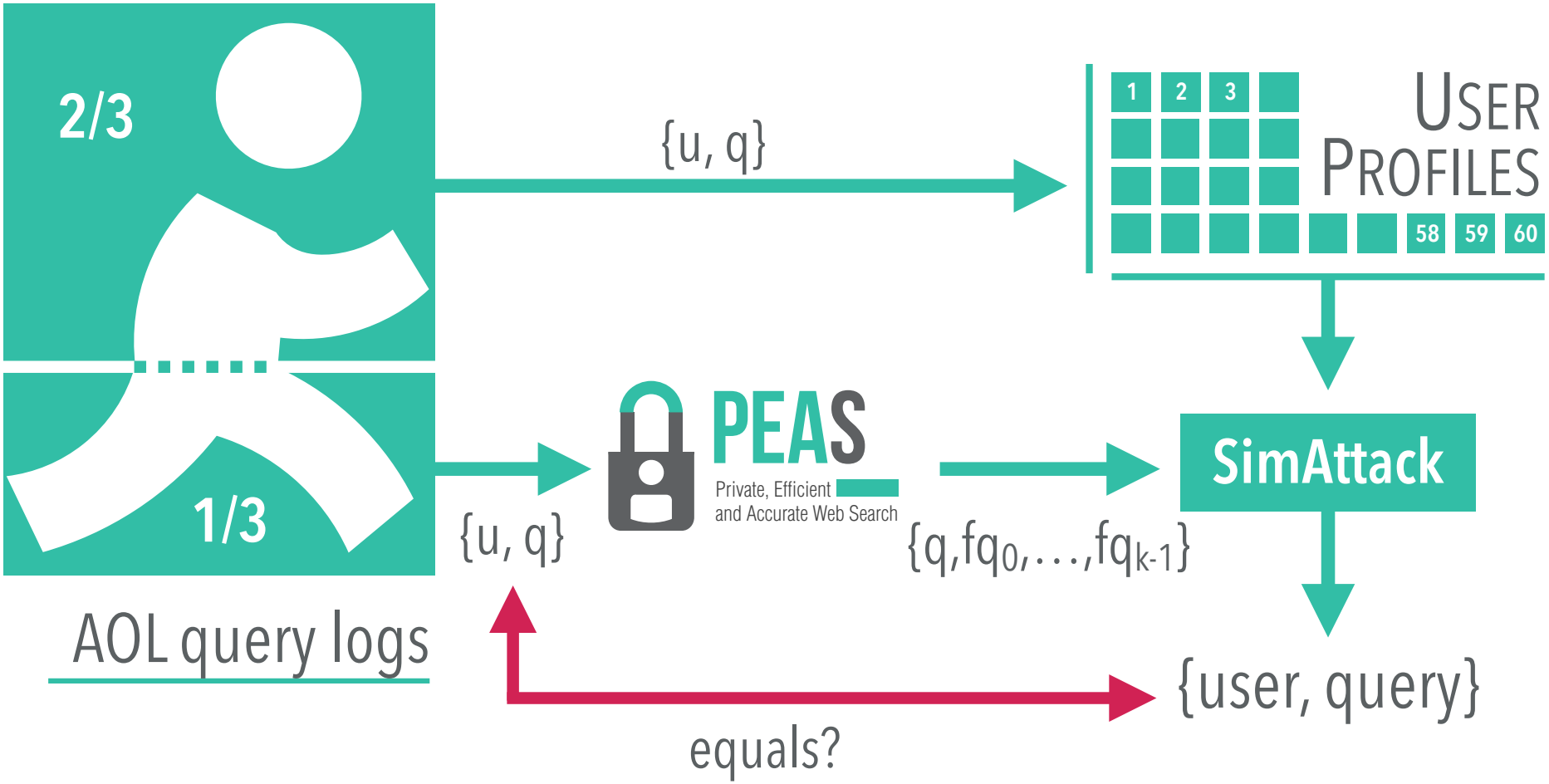
# Unlink. + Indisting.: PEAS




| | |
|---|---|
| $E(m)$ | RSA encryption of message $m$ with the public key of the issuer |
| $\{m\}_i$ | AES encryption of message $m$ with key $K_i$ |
| $Q_i$ | i-th query of user $U$ |
| $K_i$ | AES encryption key associated with query $Q_i$ |
| $A_i$ | Answer to query $Q_i$ |
| $X$ | An anonymous identifier |

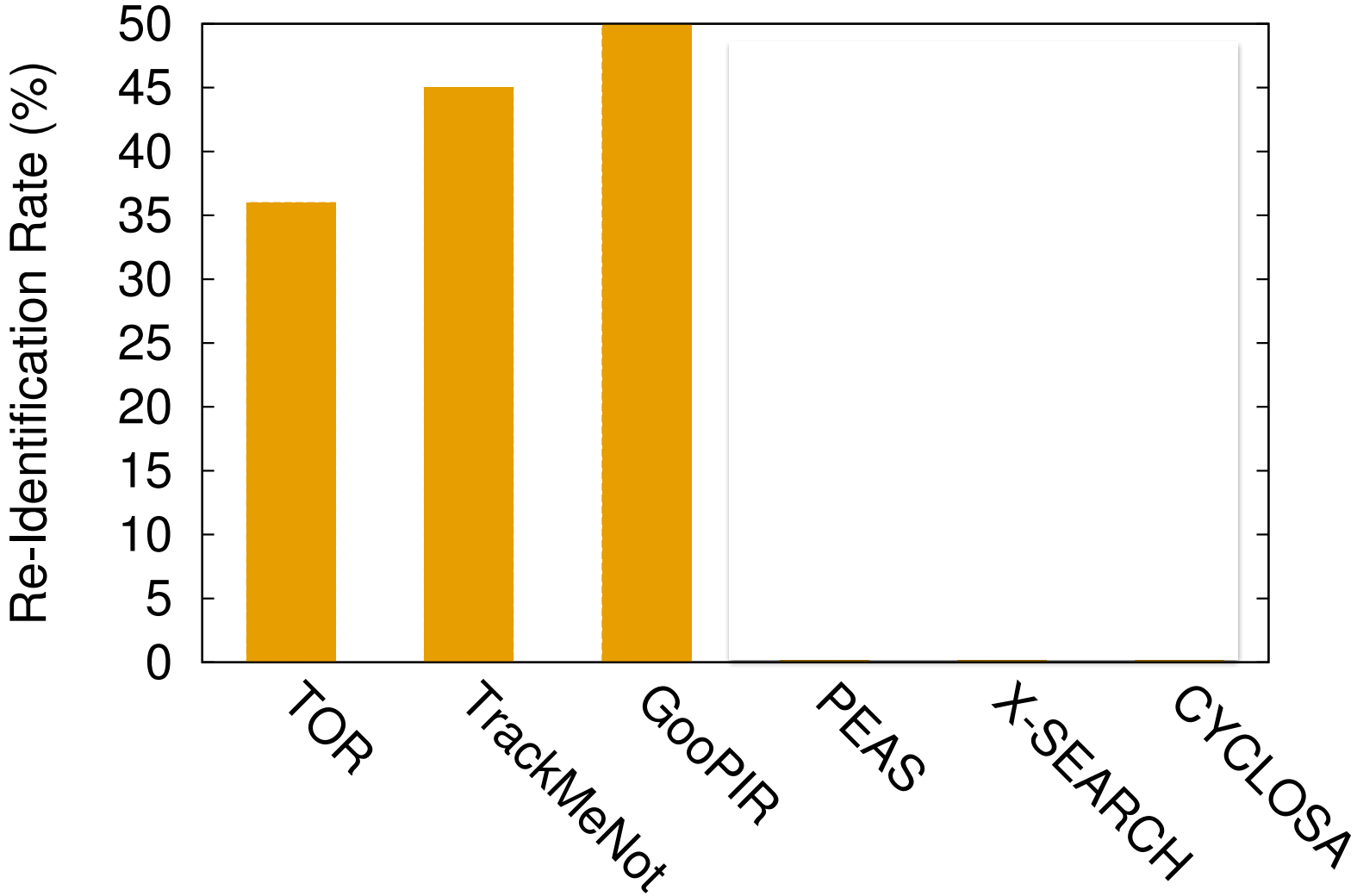
PEAS: Private, Efficient and Accurate Web Search. IEEE TrustCom'15. 2015.

# Measuring Privacy

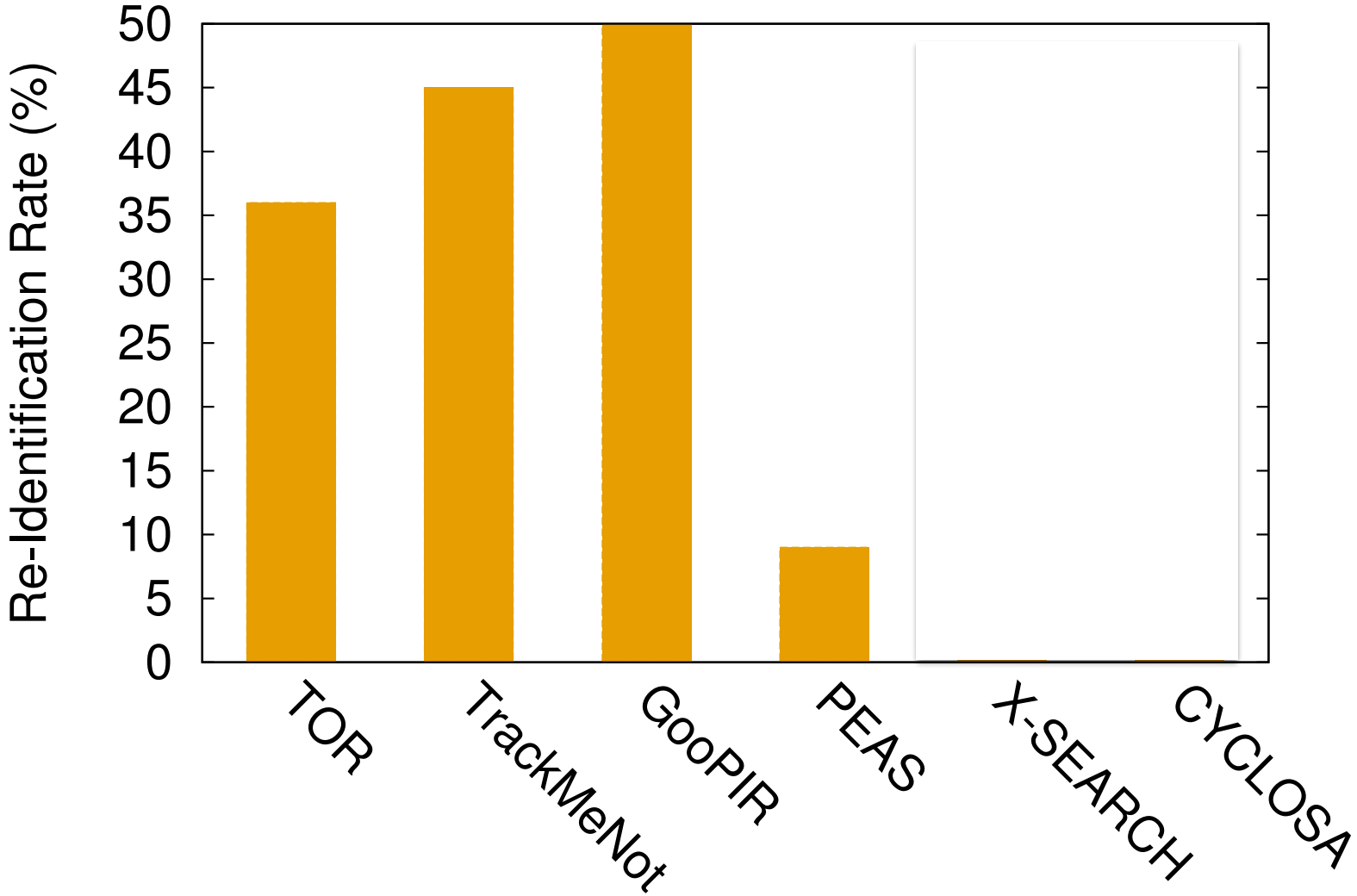



SimAttack: private web search under fire. Journal of Internet Services and Applications 7(1): 2:1-2:17 (2016).

# Measuring privacy



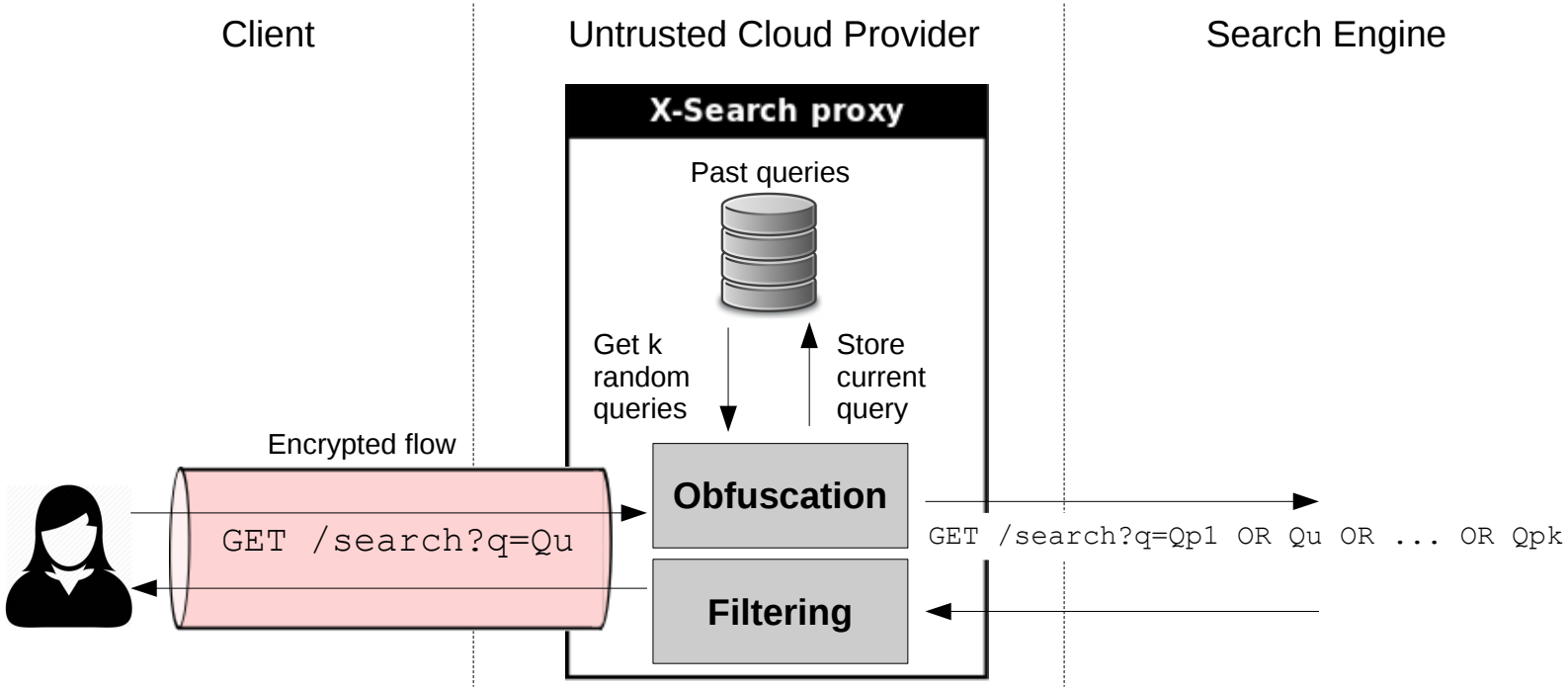
# Measuring privacy



# PEAS limitations

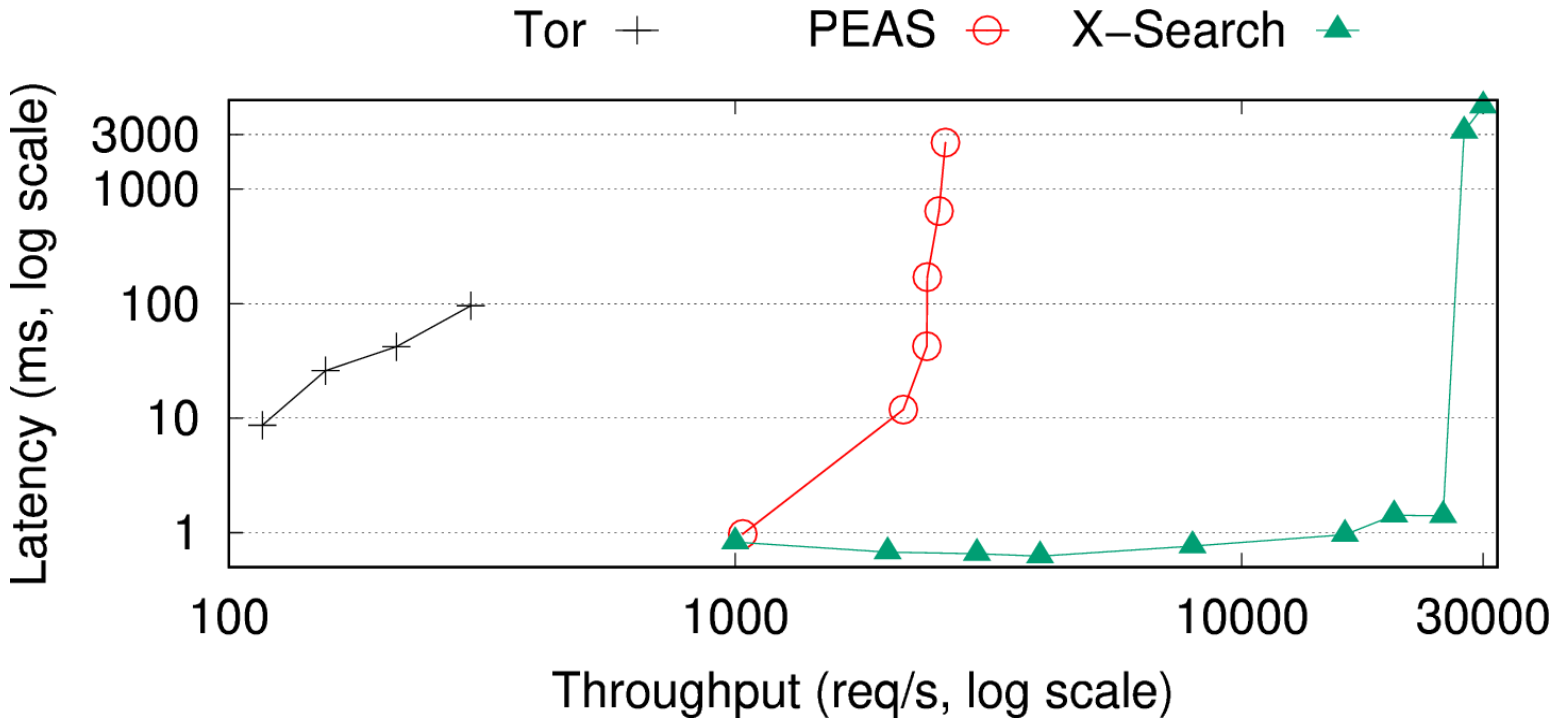
- Weak adversarial model
  - Relies on two non colluding servers
- Quality of fake queries
- Scalability

# X-Search




X-Search: Revisiting Private Web Search using Intel SGX. Middleware 2017.

# X-Search Performance

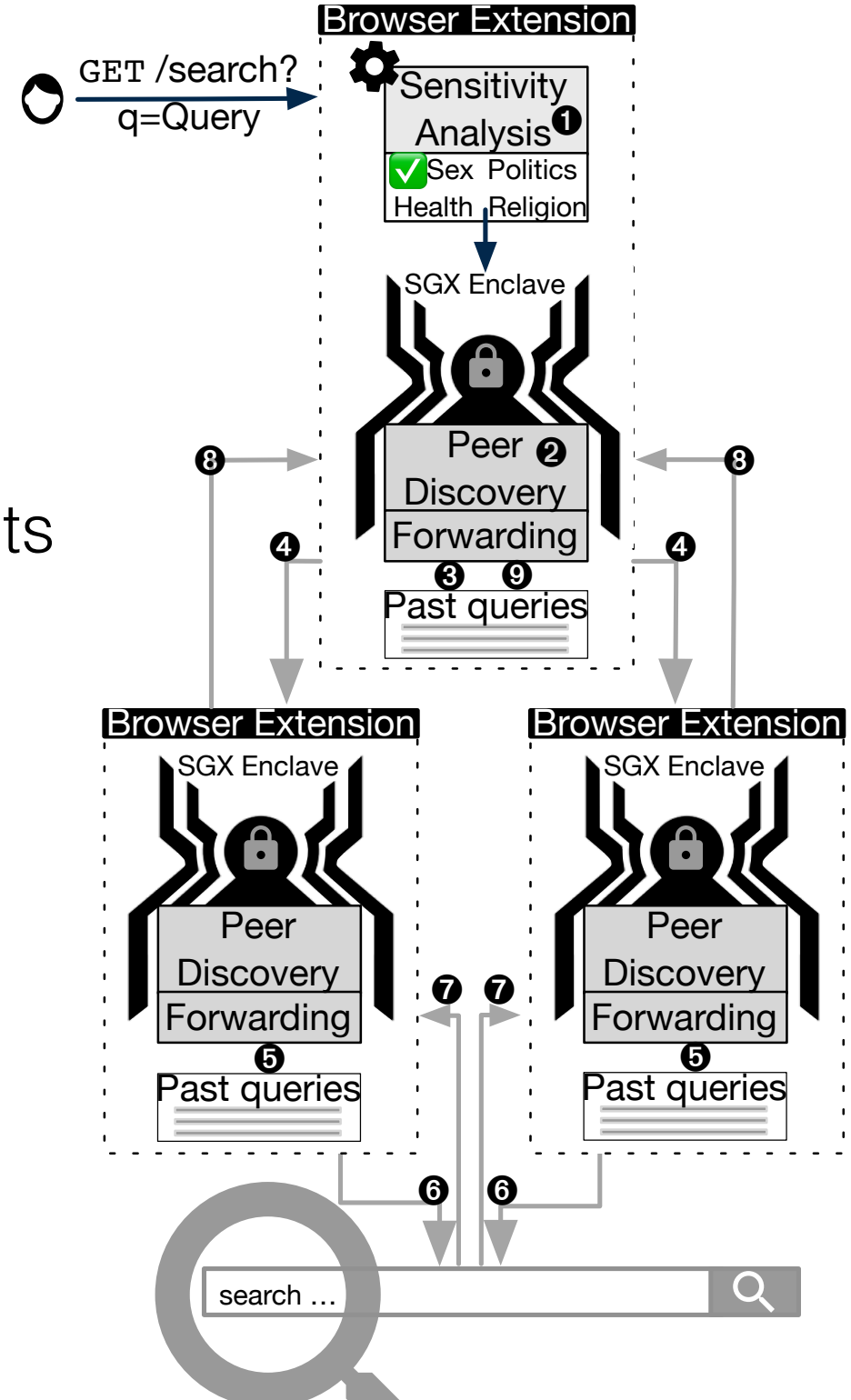


# X-Search Limitations

- Scalability
- Query limitation wrt search engine
- Accuracy

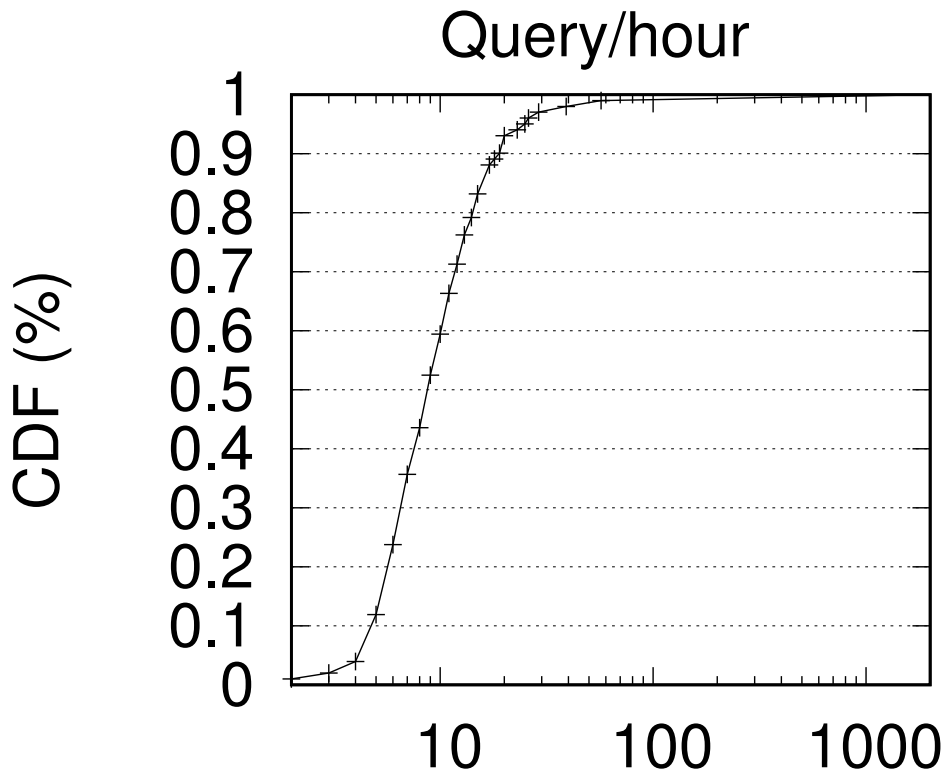
# Cyclosa

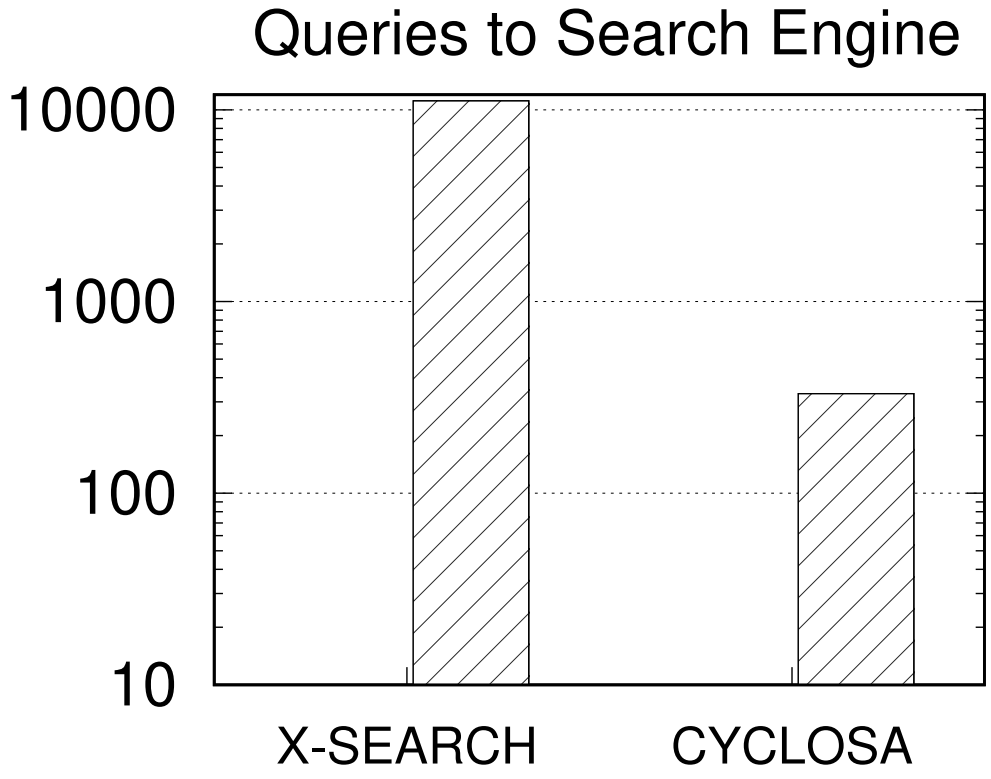
- Every node in the system acts as a proxy node for others
- Use Intel SGX
- Built as a browser extension
- Considers query sensitivity



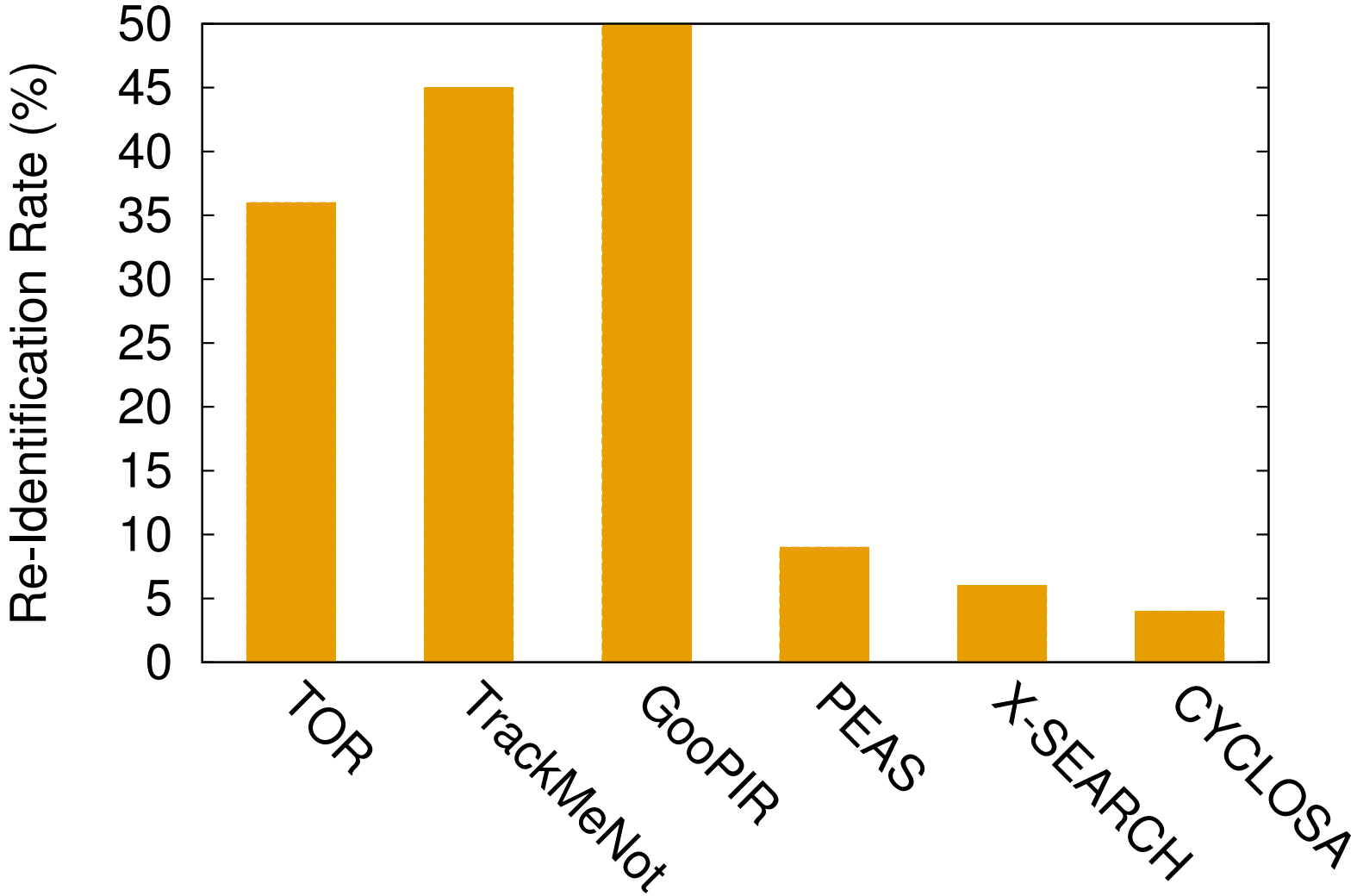

Under submission

# Cyclosa Performance



# Measuring privacy



# Sum up

- Enforcing privacy in online services is important
- Classical techniques
  - Theoretical properties vs practical privacy metrics
  - Client-side, server-side, proxy-based

# Research directions

- Generic frameworks for evaluating privacy and data/service utility
- Privacy and utility metrics
- User-centric privacy
- Address time specificities
- Adversarial ML