University of Stuttgart

Institute of Parallel and
Distributed Systems (IPVS)

Universitätsstraße 38
D-70569 Stuttgart

Understanding Vulnerabilities of Location Privacy Mechanisms against Mobility Prediction Attacks

Zohaib Riaz, Frank Dürr, Kurt Rothermel

International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous 2017)

9th Nov 2017

Background and Motivation

- Mobile apps promote location information sharing
  - Let your friends know where your are!
  - Tag tweets/photos with your location!
  - Get location-based services, e.g., nearby POIs

FOURSQUARE

- A single location update may convey:
  - Geo-location
  - Location semantics, e.g., restaurants, shops etc.

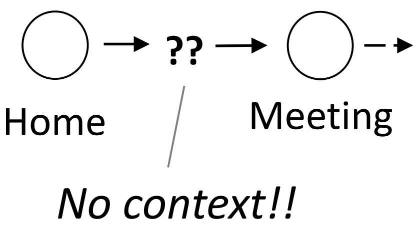
```
{(lat,lon),
venue name,
type/semantics}
```

- Sensitive semantics (e.g., hospitals)

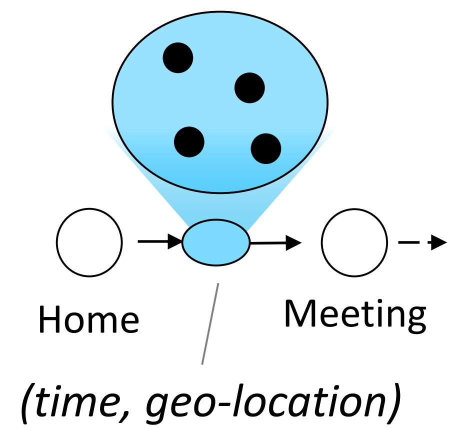
  → *User privacy concerns!*

Location Privacy mechanisms: State-of-the-art

- **Suppression:** avoids release of sensitive semantic info (Götz et al. 2012)
  - ✓ Leaks no context
  - ✓ Secure against location-history based attacks
  - ✗ cuts utility harshly: **no data → no service/sharing**

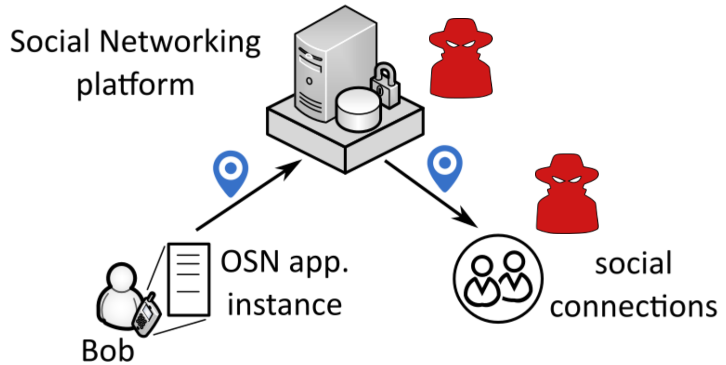


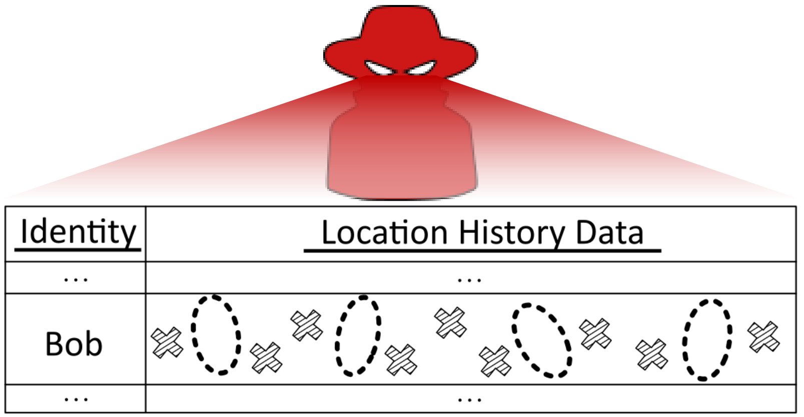
- **Obfuscation:** "cloaks" semantic info (Yigitoglu et al. 2012)
  - ✓ allows approximate POI-searches/location-sharing
  - ✗ leaks contextual information



Are existing obfuscation algorithms secure?

- **Threat Model:**
  - Attackers can aggregate location history information
    - At least in the obfuscated form
    - May possess accurate historic data





- **Hypothesis:** User privacy is at risk!

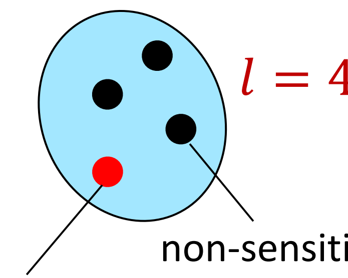
Contributions

- Design of a ***semantic mobility model*** to represent attacker knowledge
  - Both accurate and obfuscated historic location information
- Demonstration of its ***effectiveness as an attack*** against state-of-the-art semantic obfuscation mechanisms
  - Dataset: year-long location check-in histories of 278 Foursquare users
- Identification of ***fundamental design improvements*** for future semantic obfuscation mechanisms.

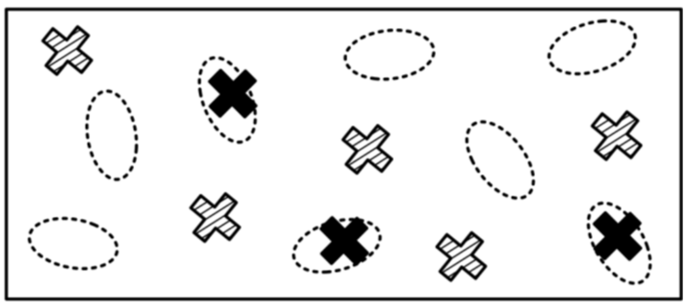
State-of-the-art Semantic Obfuscation Mechanisms

(Damiani et al. 2010, Yigitoglu et al. 2012)

- Each user can specify his sensitive semantic locations
  - *Hospitals, bars, churches etc.*
- Can also specify the degree of protection *(see paper for details)*
  - *l-diversity:* Number of distinct locations inside a cloaking region

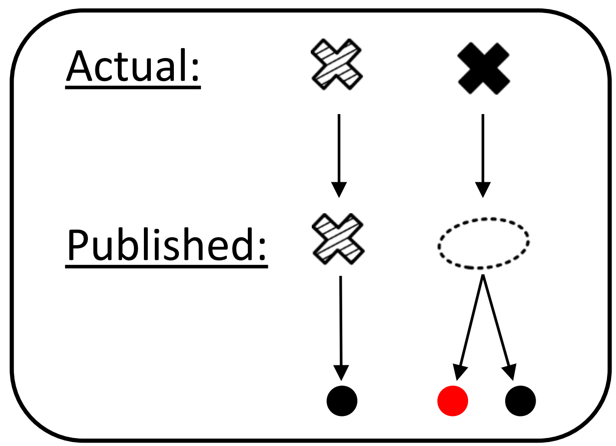


Step 1: Preprocessing

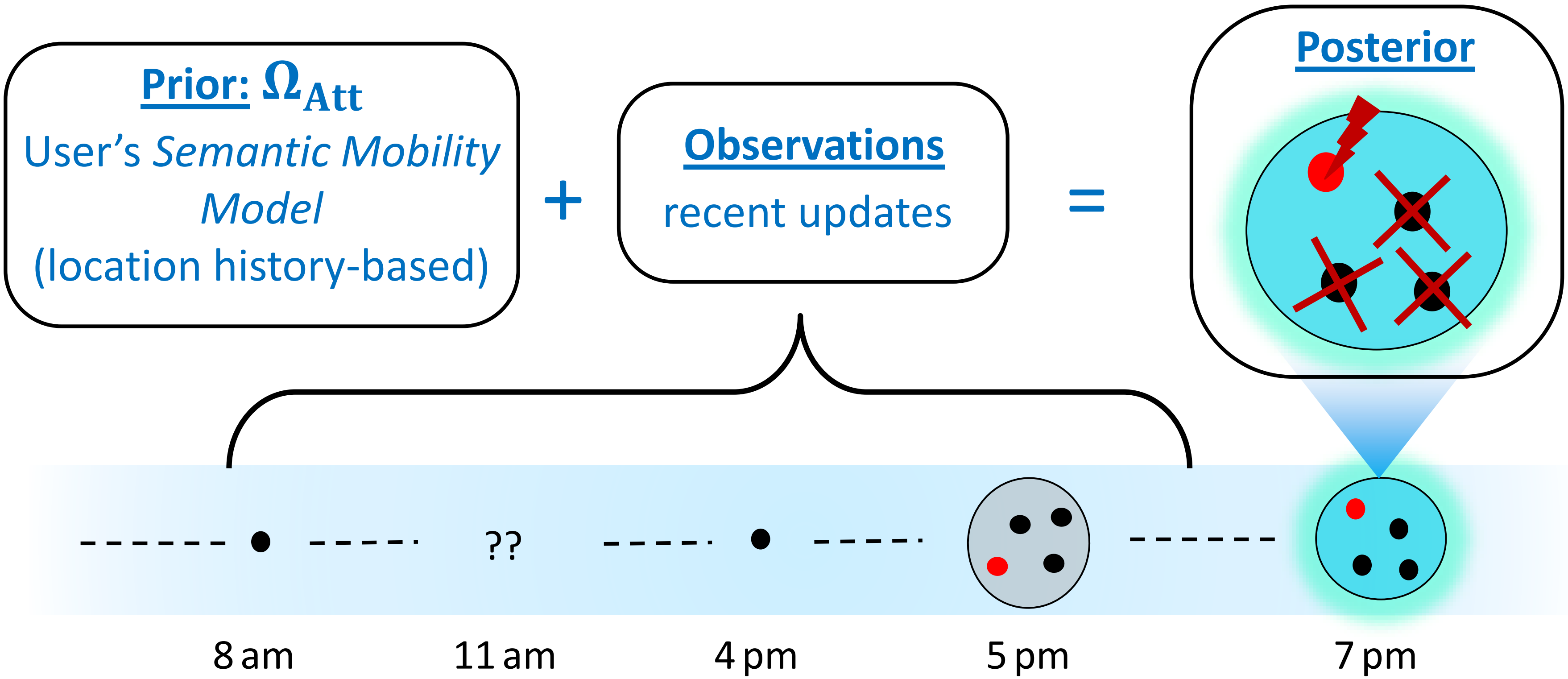


Generate CRs for City Map
(independent of user mobility)

Step 2: Real-time location updates



Attack overview



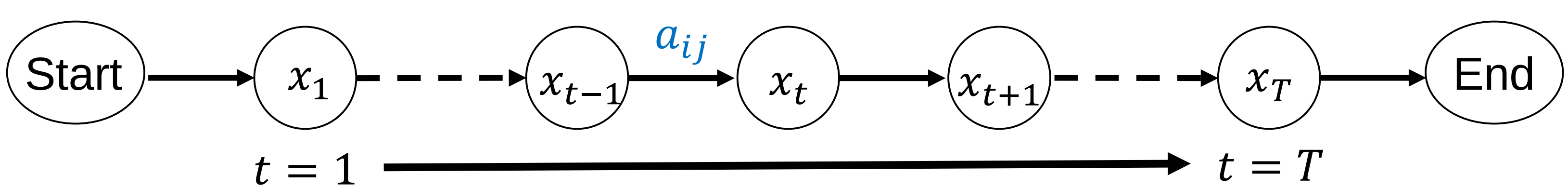
location updates in a single day …

The semantic mobility modeling problem (1)

- **Goal:** Learn $\boldsymbol{\Omega_{\mathrm{Att}}}$ from location history
- A popular fundamental assumption:

→ **Human Mobility can be modelled as a Discrete-time Markov chain**

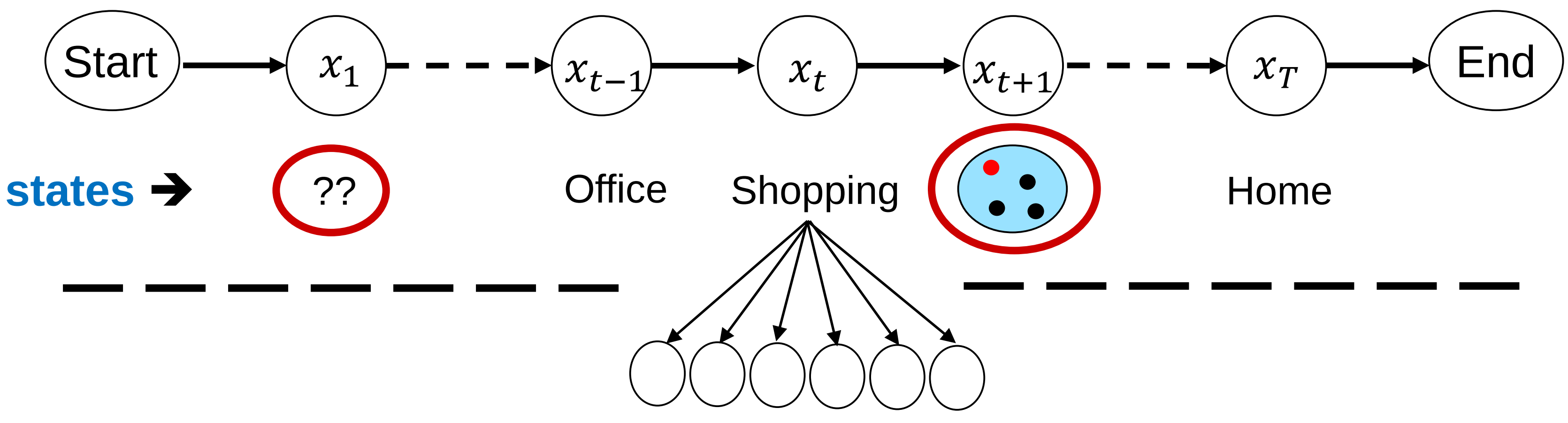
- Semantic locations modeled as **states ($\boldsymbol{x_t}$)**
  - $x_i \in S = \{s_1, \ldots, s_M\}$, e.g., home, work, shopping
- Inter-state transitions governed by probabilities:
  - $a_{ij} = P(x_t = shopping \mid x_{t-1} = home)$



Markov chain over a day of user's movement

The semantic mobility modeling problem (2)

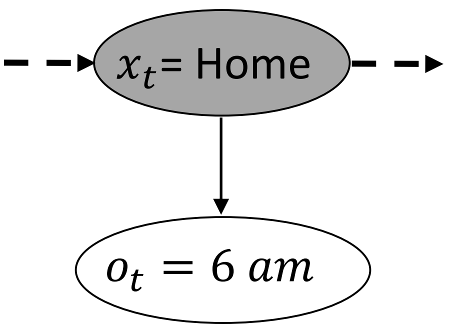
- State information:
  - not always clear
  - is additionally accompanied by other observations
- **How to model this information??**



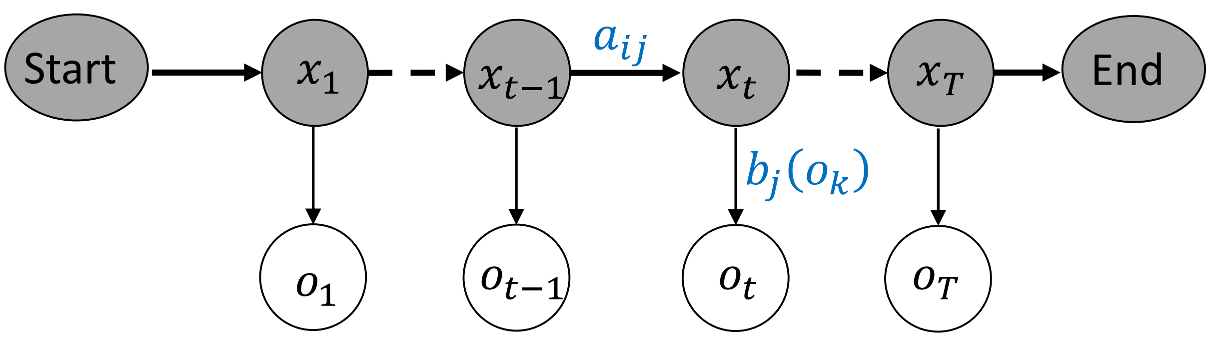
observed features ➔ {geo-location, hour-of-day, weekday, …}

Hidden Markov Models (HMMs)

- Can accommodate:
  - Cloaking Regions and missing state information
    Hidden States → $a_{ij} = P(x_t = j \mid x_{t-1} = i)$
  - Observed features as **state-dependent emissions**
    → $b_j(o_k) = P(o_t = o_k \mid x_t = j)$



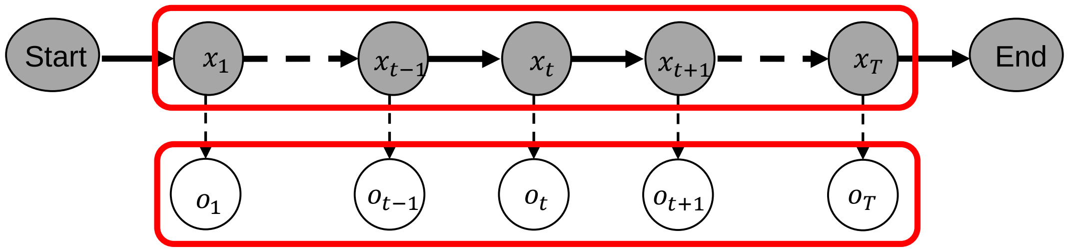
- Given $\Omega = \{A, B\}$ and $O = \{o_1, \ldots, o_T\}$
  - Can efficiently compute $P(x_t = j)$



Observation sequence

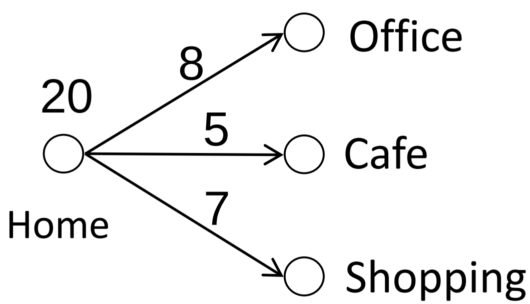
Training HMMs: Learning A and B

- **Option 1:** if dataset is **fully labeled** → *Maximum-likelihood*

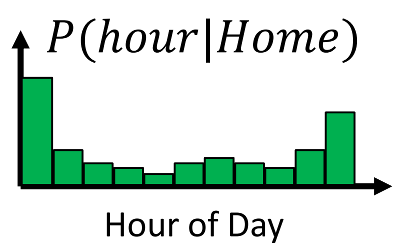


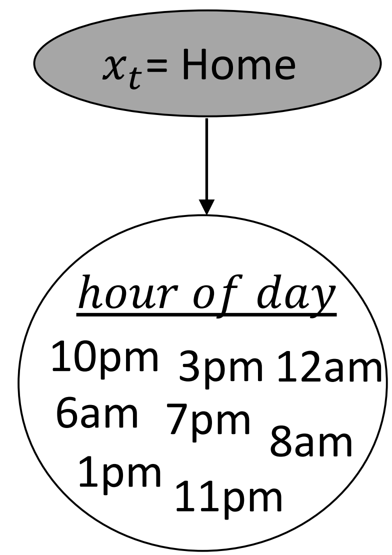
Transition probabilities

$$P(x_t = Office \mid x_{t-1} = Home)$$



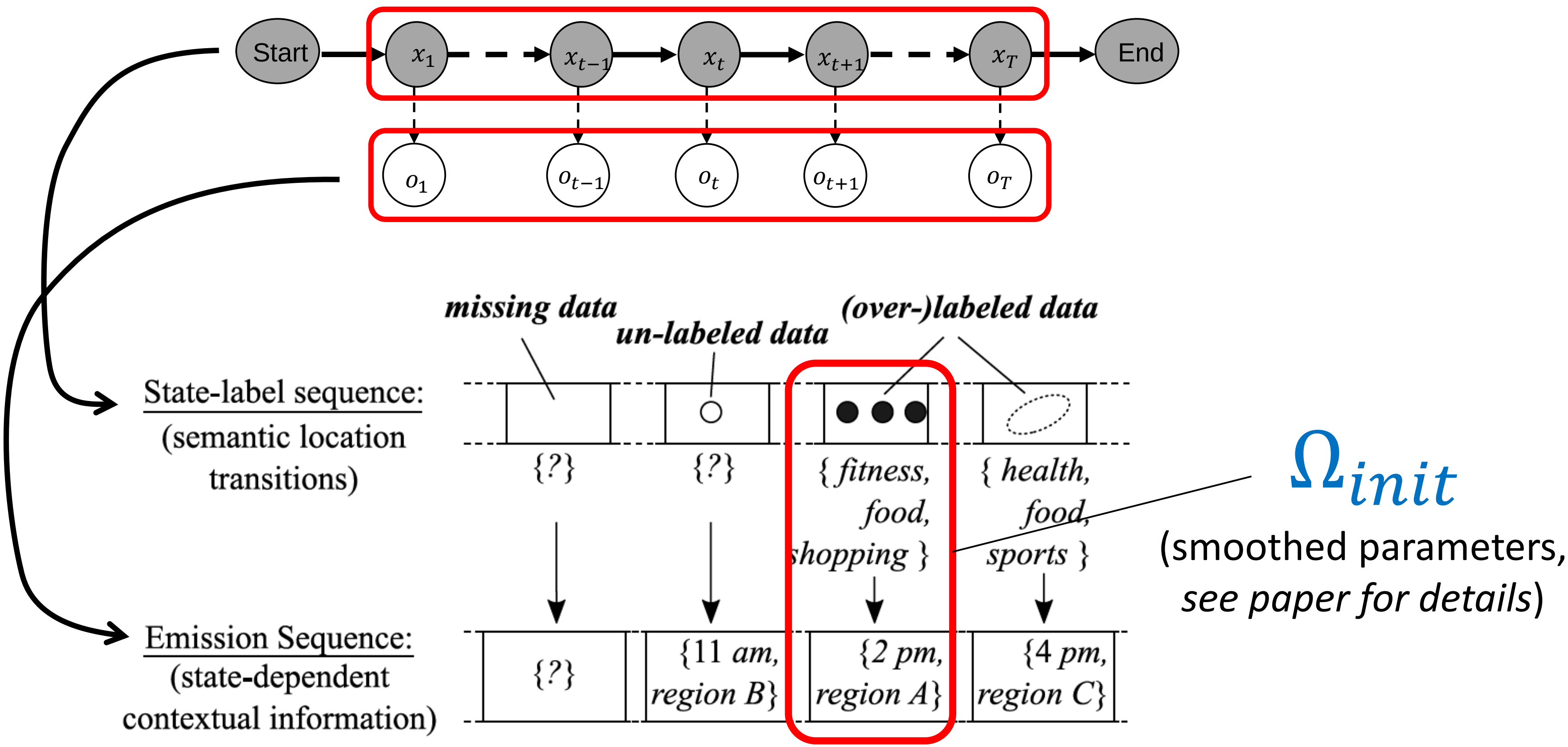
Emission probabilities





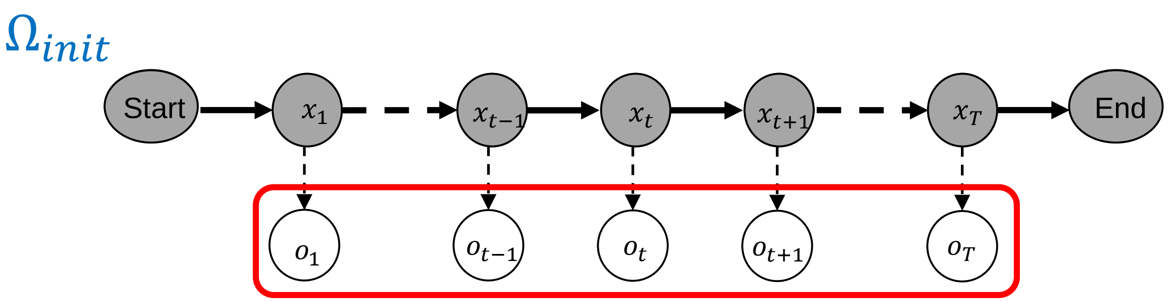
Training HMMs: Learning A and B

- **Option 1:** if dataset is fully labeled → *Maximum-likelihood*



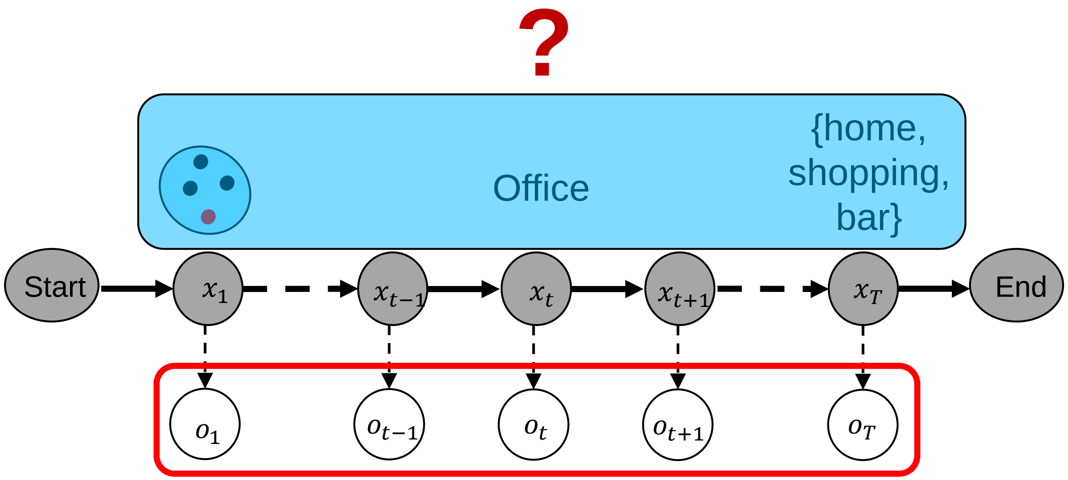
Training HMMs: Learning A and B

- **Option 2:** If labels are not present → *Baum-Welch algorithm*

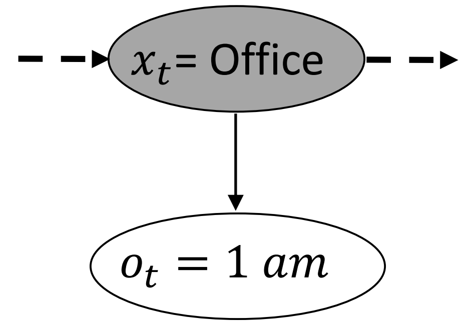


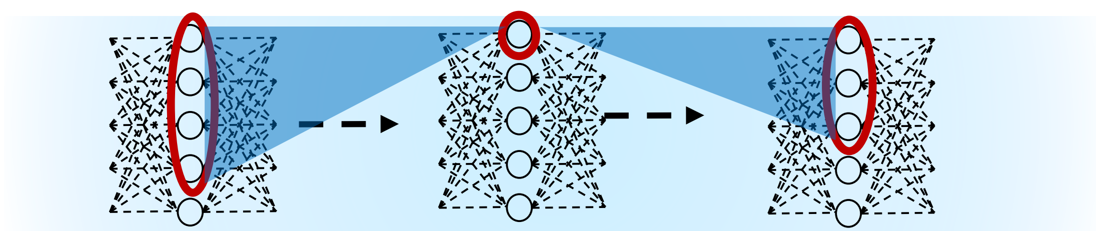
- Begin with a prior model, e.g., $\Omega = \Omega_{init}$
  - <u>Step 1</u>: generate probabilistic state-sequences using a model Ω
  - <u>Step 2</u>: Estimate Ω_{new} using *Maximum-likelihood* estimation
  - Set $\Omega = \Omega_{\text{new}}$ and repeat steps 1&2 → until convergence

Training HMMs: Learning A and B

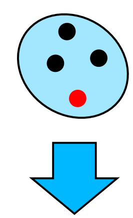


$$\boxed{P(x_t = j) \propto P(o_t | x_t = j)}$$





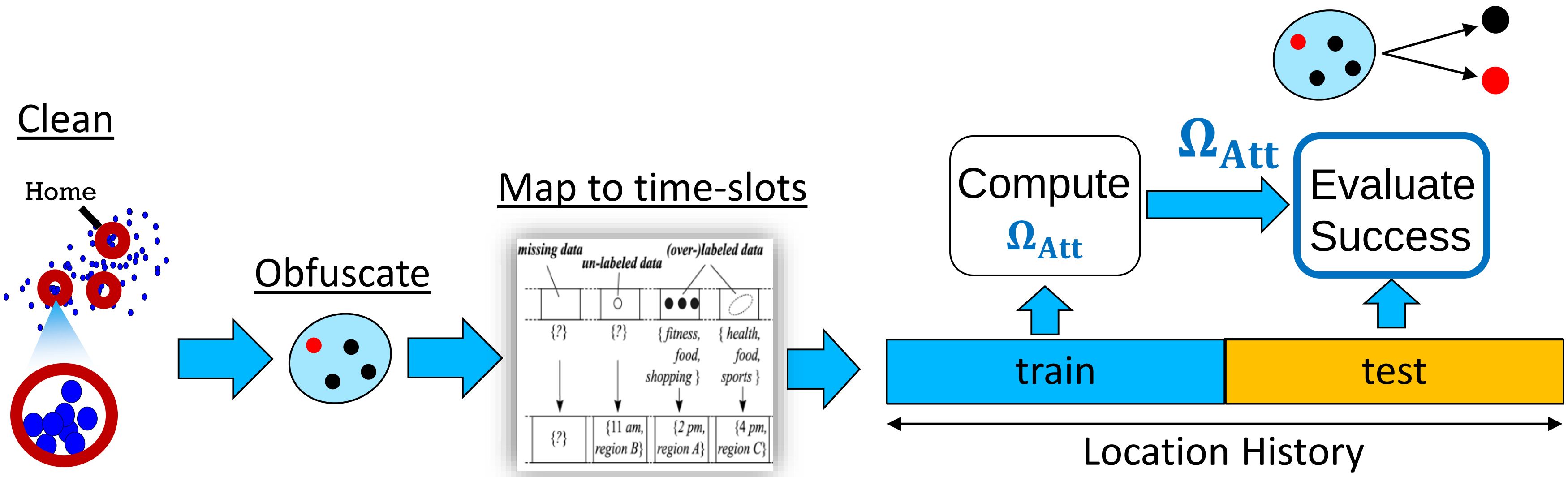
Example



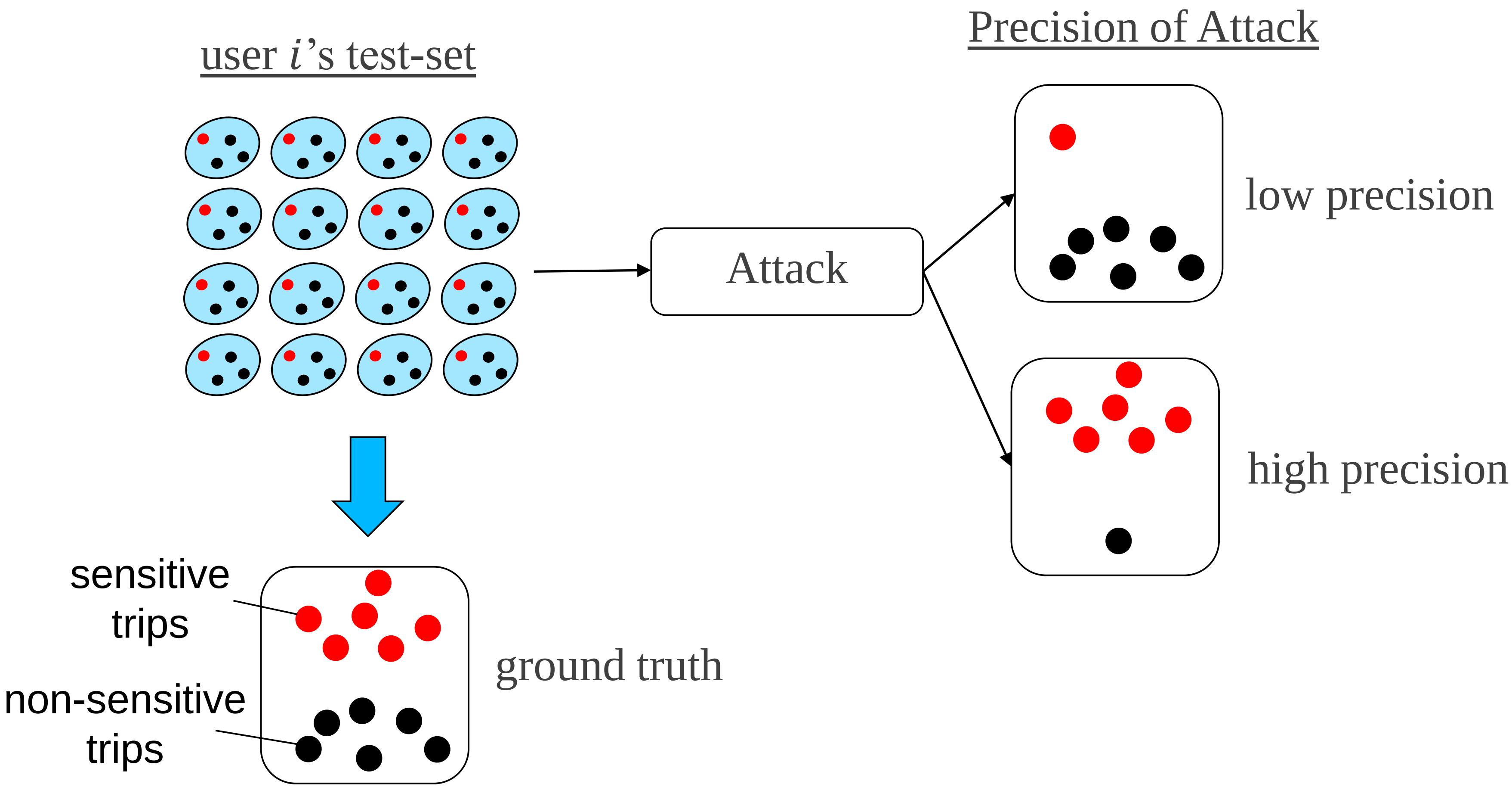
$$\widehat{b_j}(o_t) = 0 \ \forall \ s_j \notin CR$$

Experimental Workflow

- **Dataset:**
  - Crawled check-ins from Twitter’s public feed from Nov 2015- Nov 2016
  - Got venue information (including surrounding venues) from Foursquare
  - Necessary filtering leaves **278 users** with **284,472 check-ins**
  - Mean length of check-in history: **246 days**

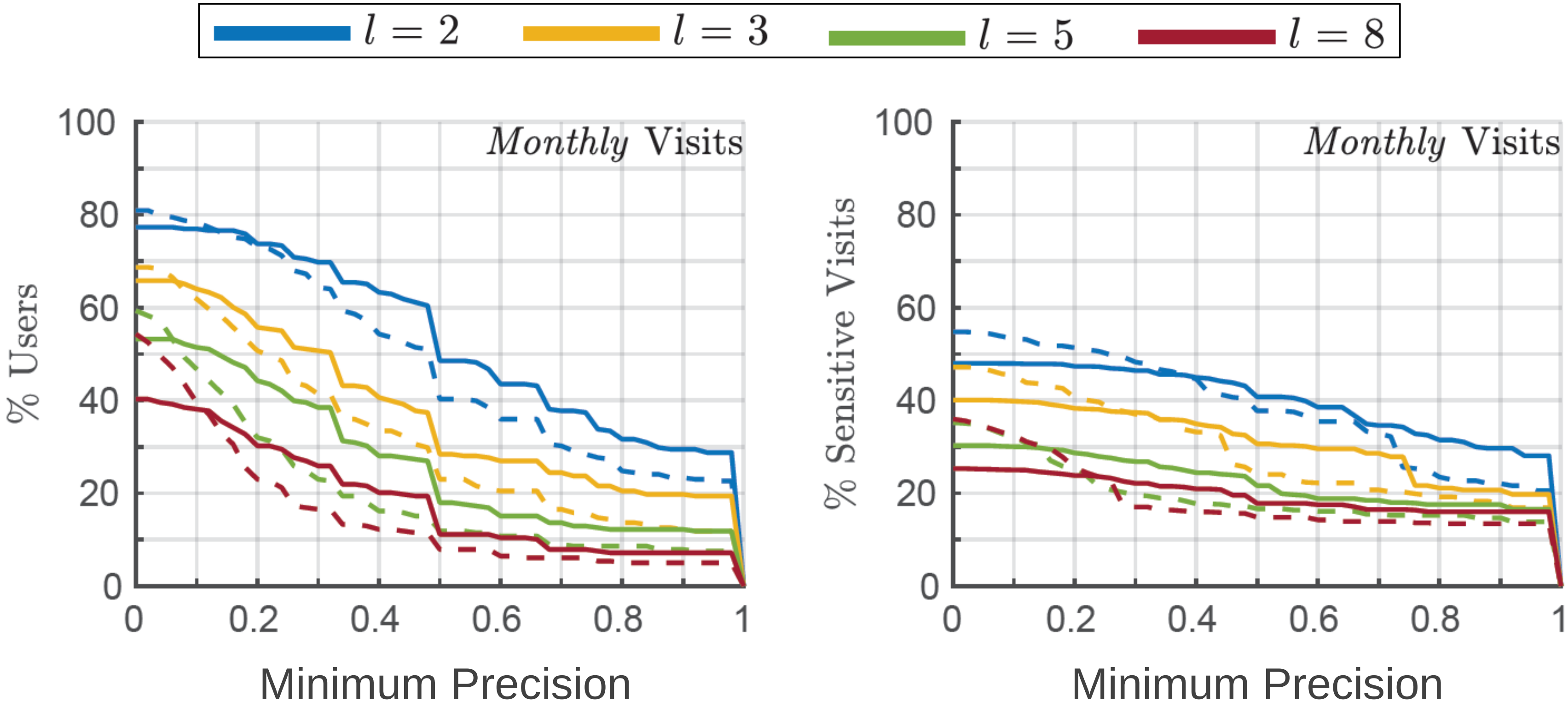


Evaluation Metrics



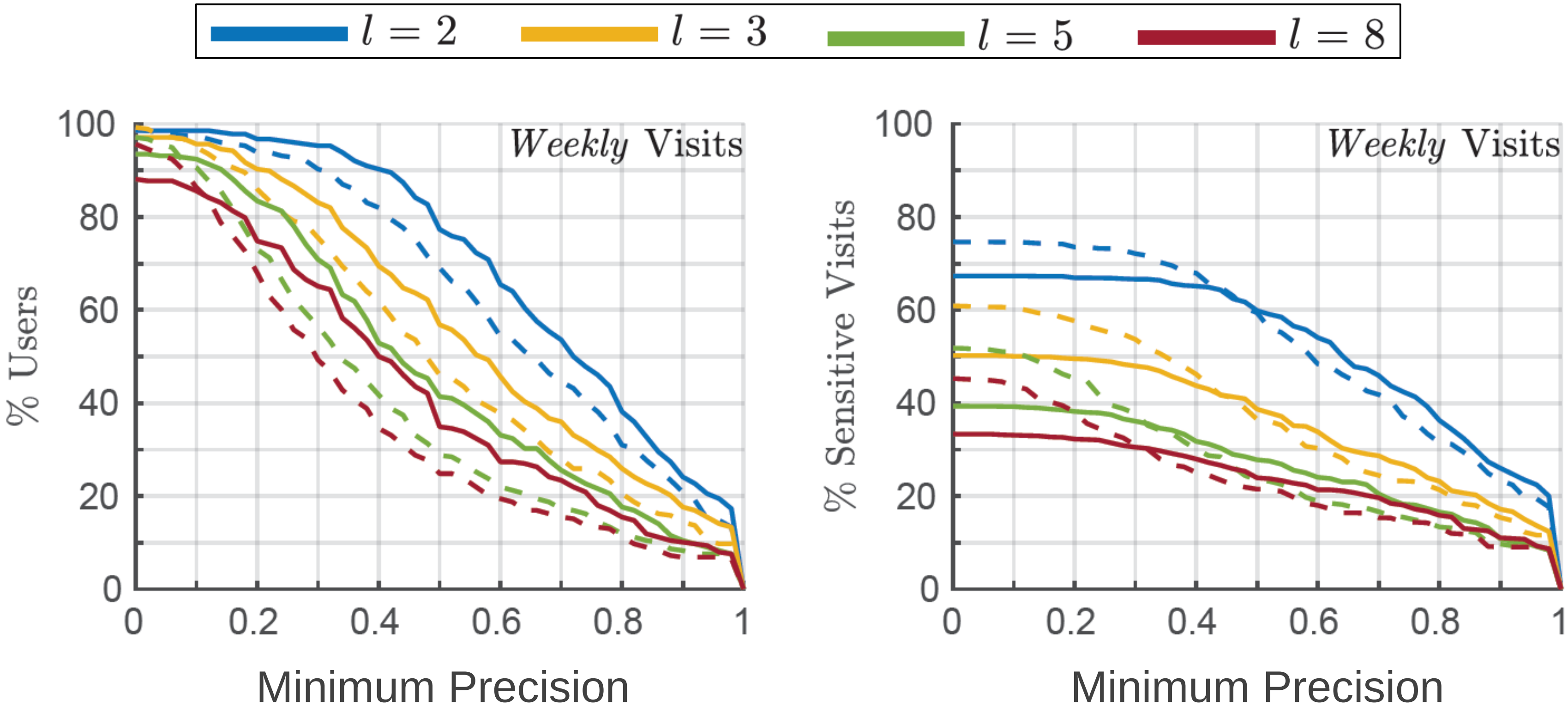
Results

- *Sensitive location selected s.t. it is visited at a certain frequency*



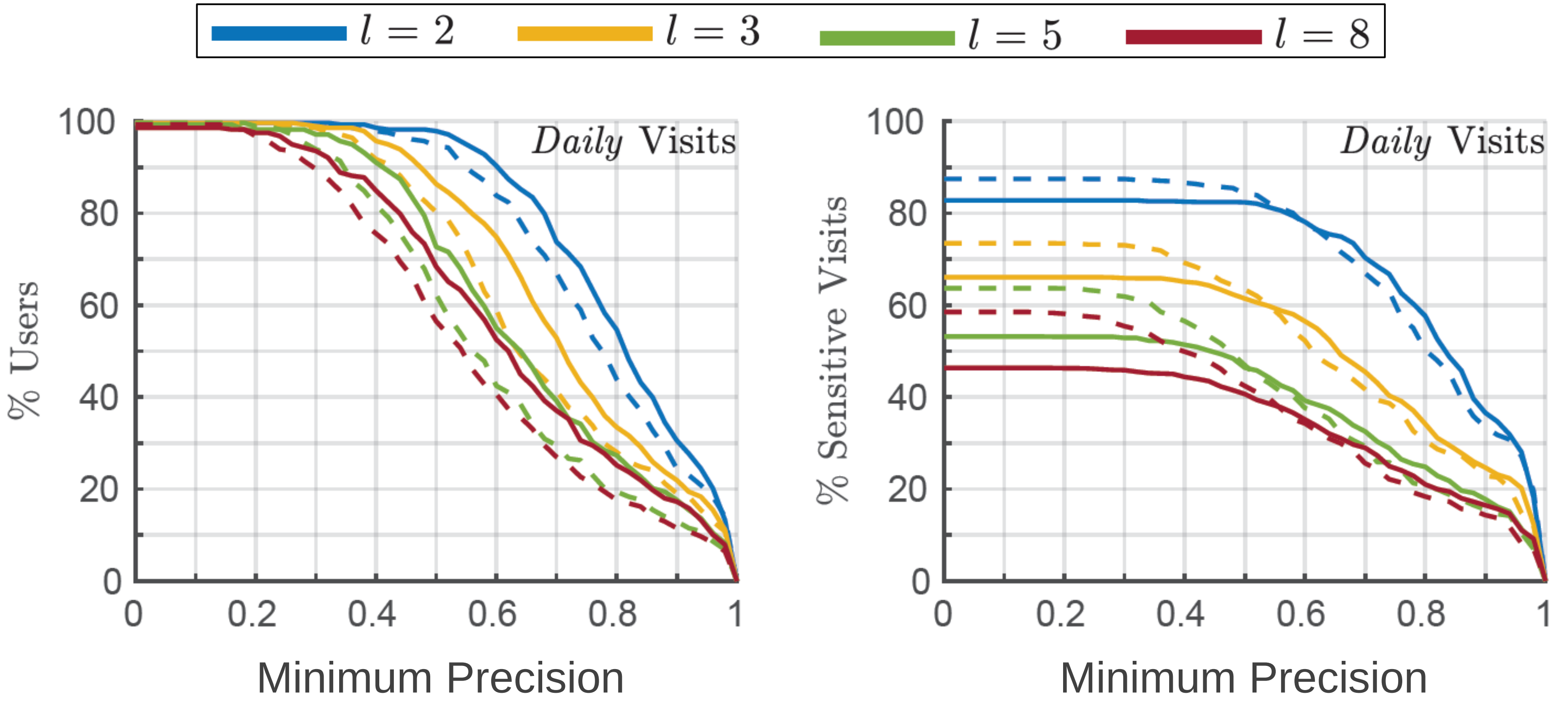
Results

- *Sensitive location selected s.t. it is visited at a certain frequency*



Results

- *Sensitive location selected s.t. it is visited at a certain frequency*



Conclusion

- We show that the **privacy guarantees** offered by state-of-the-art location obfuscation mechanisms **are weak!**

- Obfuscated location-history
  - can be exploited for mobility modeling
  - can be used to de-obfuscate user trips

- State-of-the-art location obfuscation mechanisms are **more vulnerable to de-obfuscation when used frequently**

- The need of **mobility-aware obfuscation algorithms** is evident!

Contact and Discussion

PriLoc

www.priloc.de

Zohaib Riaz

Institute for Parallel and Distributed Systems,

University of Stuttgart, Germany

zohaib.riaz@ipvs.uni-stuttgart.de