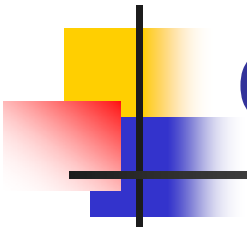
Protecting Location Privacy through Semantics-aware Obfuscation Techniques

Maria L. Damiani^, Elisa Bertino*, Claudio Silvestri^

^Dept. Computer Science, University of Milan (I)

*Dept. Computer Science, Purdue University (USA)



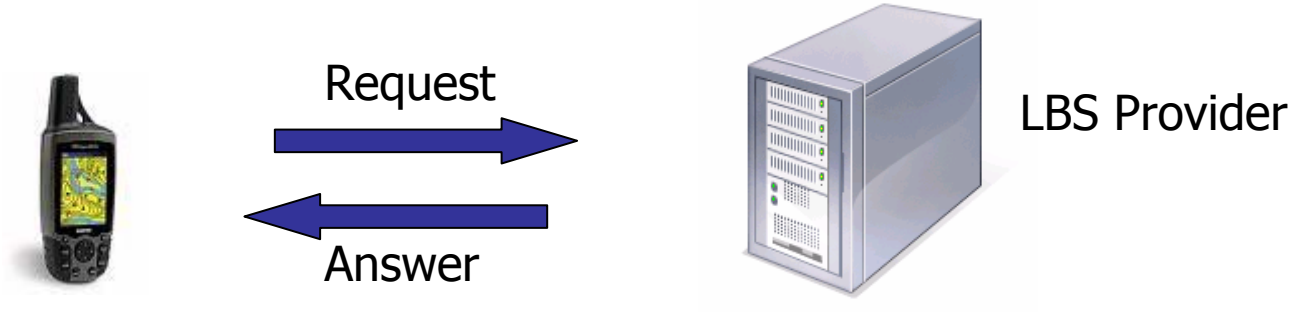
Contents

- Location privacy in LBS
- The privacy attack
- The privacy-preserving strategy
- Conclusions and future plans

The LBS context

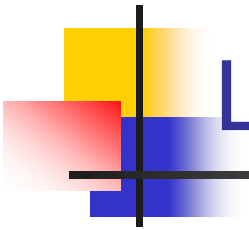
LBS is becoming a key application area under the push of GPS-enabled handset providers, novel satellite navigation systems and special events

Location privacy in LBS



Request: [U345, “Where is the closest ATM” , (long, lat)]

- Accurate position information is sensitive, LBS providers are not fully trustworthy
- The location privacy issue: how to prevent the communication of the association user-position (Beresford & al. 2003)
- Focus of this work: preventing the disclosure of a user’s position in a place which is sensitive for this user, e.g. a hospital



Location privacy-preserving techniques

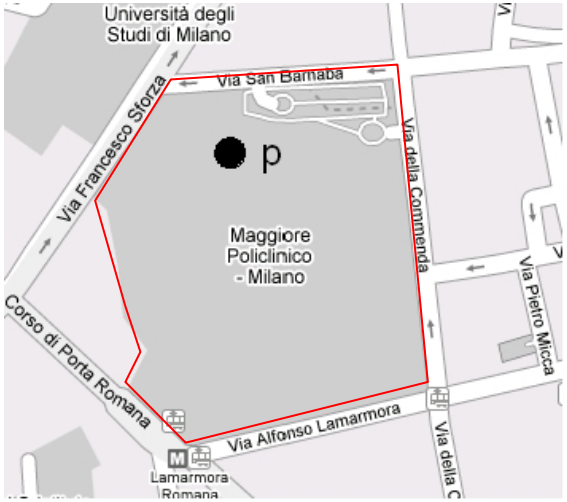
- Location k-anonymization (Gruteser03, Gedik05, Mokbel06, Kalnis07)
  - The goal is to make the request anonymous. Yet, the user's location is a quasi-identifier
  - The location is thus generalized, i.e. it is made undistinguishable from the location of other k-1 users

- Location obfuscation (Atallah04, Duckam05, Cheng06, Ardagna07, Yiu08)
  - The goal is to provide correct answers to LBS queries without knowing the position of the user
  - The client transmits a coarse or fake position along with the query; the LBS provider sends back a set of possible solutions; the client selects among the set of candidates the most suitable

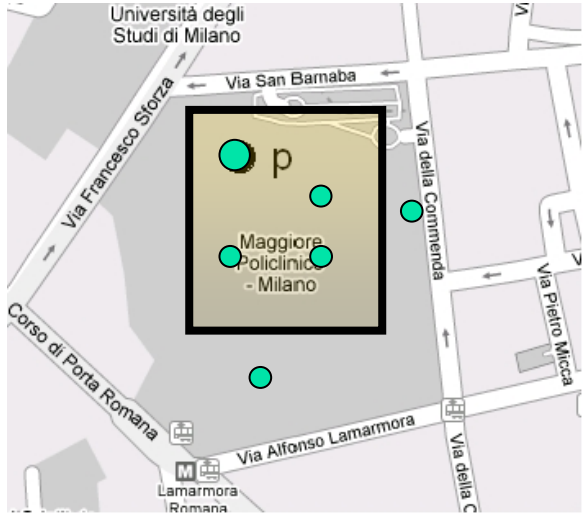
The spatial knowledge attack

- Current methods typically consider position from a geometric viewpoint while, conversely, position may have also a meaning
  - A point of coordinates (x0,y0) vs. to be in a hospital
- Current techniques are unable to protect against the inferences made by linking the geometric information with the location meaning, that, depending on the user, may represent sensitive information

Example



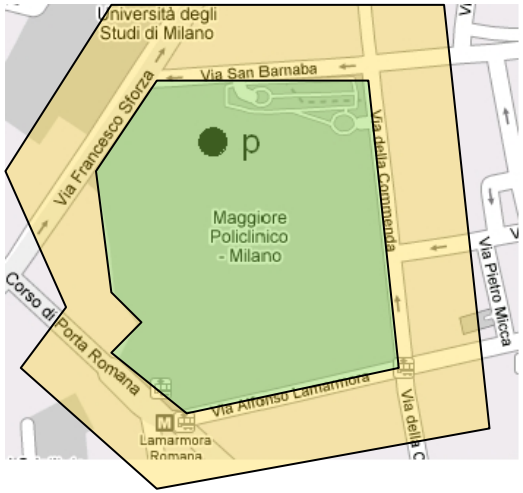
4-anonymous location



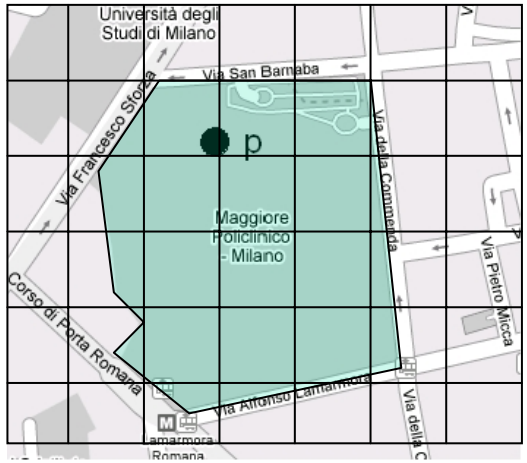
Outline of the privacy-preserving strategy

- A user specifies preferences about sensitive places and the desired degree of privacy
- The location obfuscation process is applied
  - Step 1) Off-line: obfuscated locations generation
    - For each sensitive place a coarse location is generated
  - Step 2) At run-time: obfuscation enforcement
    - The user's position *p* is matched against the set of obfuscated locations. If *p* is contained in any obfuscated location *L* then *L* is transmitted to the LBS provider
- An adversary who only knows the obfuscated location may only infer that the user *may be* in a sensitive place

The obfuscation of regions



A naive approach



Each cell is assigned a sensitivity value

Each cell is obfuscated separately

Obfuscation is obtained by merging adjacent cells, until the desired degree of privacy is possibly obtained

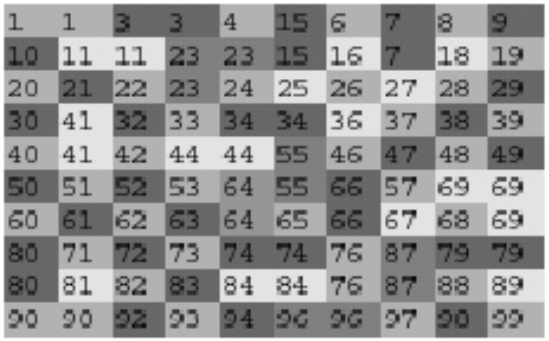
The specification of privacy preferences

- Model of space
  - Sensitive and unreachable location classes
    - S*ensitive* locations classes: Religious Buildings
    - *Unreachable* locations classes: Military Zones
- Sensitivity metric
  - One of the possible metrics:
    - Each class of sensitive locations is assigned a score
    - $SL(r) = \sum_{ft \in FTs} Score(ft) \ Area_{Fea}(r,ft) / Area_{Reg}(r)$
- Sensitivity threshold θ_{Sen}
    - Specifies the maximum sensitivity tolerated by the user
    - $\forall r \ SL(r) \leq \theta_{Sen}$

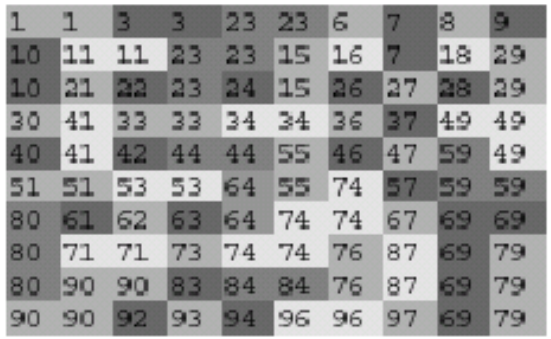
The obfuscation algorithm: issues

- 1) How to represent and how to aggregate the cells?
  - Regions are represented as nodes of a graph (RAG - Region Adjacency Graph). Arcs connect adjacent regions.
  - Cells are merged and larger regions are obtained by shrinking the graph
- 2) On which data can we evaluate the algorithm?
  - Available data sets describe locations at low resolution. A generator of synthetic data has been developed for the population of areal sensitive place

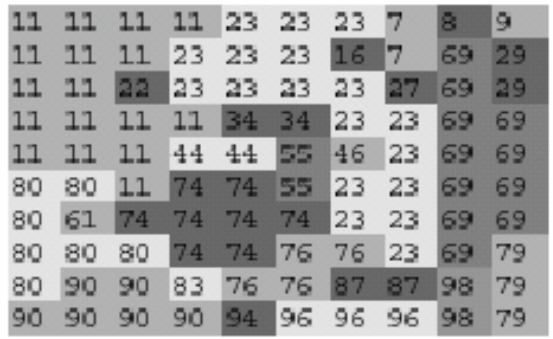
Experimenting with synthetic data



$\theta_{sens} = 0.5$



$\theta_{sens} = 0.4$



$\theta_{sens} = 0.3$

Final remarks and future plans

- We have identified and then proposed an approach to contrast the spatial knowledge attack under the assumption that all positions in space are equally probable
- Focus on: privacy model, obfuscation algorithm
- Open issues:
  - Generalization to non-uniform distribution of positions
  - Evaluation of more intuitive sensitivity metrics
  - Evaluation of different heuristics, to ensure a scalable solution
  - Protection from additional context-based inferences