Analyzing semantic locations cloaking techniques in a probabilistic grid-based map

[Demo paper]

Maria Luisa Damiani
University of Milan, Italy
damiani@dico.unimi.it

Claudio Silvestri
University of Venice, Italy
silvestri@dsi.unive.it

Elisa Bertino
Purdue University, USA
bertino@cs.purdue.edu

ABSTRACT

The protection of semantic positions, for instance that an individual is inside a hospital, is a challenging privacy issue. For example, it can be shown that popular protection mechanisms, like location cloaking, can be easily defeated when certain mobility patterns are known, e.g., certain places are more or less popular than other places. To prevent this kind of attack, novel semantic location cloaking heuristics are being investigated. These methods are built on the knowledge of population distribution and sensitive locations. In this demonstration, we present SAWL (Semantics-aware Location cloaking), a tool supporting the comparison of semantic location cloaking methods over real and synthetic spatial scenarios.

Categories and Subject Descriptors

H.2.8 [**Database management**]: Database applications—*Spatial databases and GIS*; K.4.1 [**Computers and society**]: Public Policy Issues—*Privacy*

General Terms

Design, algorithms

Keywords

Privacy, location-based services, mobility models

1. INTRODUCTION

Geolocalization services are becoming dramatically pervasive in mobile Internet pushed by the advances in positioning technologies (e.g., hybrid positioning systems), emerging standards (e.g., W3C Geolocation API/ HTML 5) and the growth of the mobile advertising market. These trends pave the way to the collection of huge amounts of mobility data from which mobility patterns, e.g., how people move and where socialize, can be extracted. In such a setting, privacy is a major concern.

Location cloaking is a widespread used privacy protection technique. It simply replaces the exact user's position p with a cloaked region (CR) p' containing p. CRs can be used to protect both the user's *identity* and the *location privacy* against inference channels. However, location cloaking does not offer adequate protection when the places that people frequent and the degree of frequentation are known. For example, if a CR includes a highly frequented hospital and a desolate park adjacent to the hospital, an observer with such information and knowledge about population distribution can promptly prune the CR and infer that the user is very likely in the hospital (and thus may have health concerns).

Against this inference, novel techniques supporting *semantic location cloaking* have been recently investigated [1]. These methods are built on the assumption that the space is partitioned in locations of different type, e.g., buildings, streets and so forth, and the distribution of population in space is known. In such a context, the CRs are generated so as to bound the probability of user's assciation with sensitive locations. The cloaking service can be provided for example by a trusted location service provider: clients request their location to the location service provider (e.g., Skyhook Wireless), get a possibly cloaked location and then forward it to the application. Different cloaking methods can be envisaged to limit the loss of spatial accuracy.

SAWL is a compresive framework supporting the analysis of semantic location cloaking methods through a friendly visual interaction environment. SAWL can be used by scientists to support the development of novel cloaking solutions as well as by trusted location service providers to configure location cloaking services. This demonstration illustrates the basic features of SAWL and how it can be used.

2. THE SAWL FRAMEWORK

The space of interest is organized in a grid. The grid can be populated of sensitive locations of different type and size. Moreover, SAWL generates non-uniform population distributions over the discrete space, based on a model inspired by the Community-based Mobility Model (CMM) [2]. The idea borrowed from CMM is that individuals tend to gather in locations (called Attraction Points in our solution) which exert a different social attractivity on people. The degree of attractivity and frequentation of Attraction Points is modelled in SAWL using a stochastic approach.

SAWL provides multiple cloaking methods and supports the specification of privacy profiles. Various metrics can be utilized to measure the QoS. Additional functionalities

ACM GIS '10 San Jose, California USA

support data import/export of probability and sensitive locations grids so as to allow the interoperation with external systems and the use of real data.

2.1 The system architecture

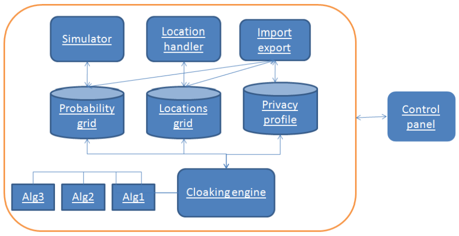


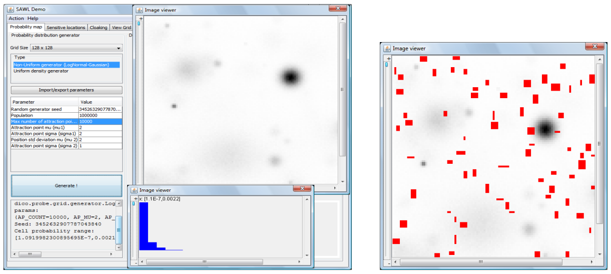
Figure 1: SAWL architecture

The SAWL framework is implemented in Java. The system architecture in Figure 1 consists of five modules. The **Control panel** handles the users input and the visualization aspects. The **Simulator** generates the probability distribution of population (termed Probability Grid, PG). A non-uniform probability distribution is generated by randomly associating each individual from a given population to an Attraction Point (AP) and then, in order to refine the position, to a cell in proximity of that AP. The PG is finally obtained by normalizing the number of observations in each cell over the whole population. Currently the associations between users and AP, and between AP and cells are modelled using a LogNormal and bivariate Gaussian distribution, respectively. The **Location Handler** populates the grid space of sensitive locations, e.g., hospitals, as rectangles of fixed or randomly varying sides. The **Import/export** module supports data exchange with external real data sources. The **Cloaking Engine** handles the cloaking algorithms and privacy profiles. Currently, there are three cloaking algorithms, called respectively $Sens_{Reg}$, $Sens_{Div}$ e $Sens_{Hilb}$. These methods are suited in different settings (we refer the reader to [1]). The outcome of each algorithm is a set of CRs and a set of QoS measures indicating: a) the average size of CRs; b) the probability-weighted sum of CRs size (i.e., the expected size of the CR); c) the average sensitivity of CRs; d) the cloaking processing time. An additional measure is the number of CRs.

3. DEMONSTRATION OVERVIEW

The demonstration illustrates how to construct synthetic scenarios and perform "what-if" analysis to explore the impact of varying privacy profiles, spatial settings and cloaking methods over QoS. The scenarios are visually represented in *Map* windows to allow a quick visual analysis.

Probability Grid construction. The grid size (e.g., 128×128 cells, assuming a cell resolution of 10 metres) and the parameters for the construction of the PG (mean and standard deviations) are set through the panel in Figure 2.a. The PG is visualized in the Map window of the same panel; by convention the darker grey areas in that map are those more populated. In addition a frequency histogram reports the number of cells for the different ranges of probability. In the example, the histogram shows that the majority of cells are not much frequented while a limited number of cells are highly populated. **Location generation**. The names of the sensitive location types, e.g. hospitals, are input together with the desired percentage of coverage and the parameters for the geometric construction. Locations are then generated and finally added to the Map window as red rectangles (Figure 2.b). The Map window of the example shows a low density of sensitive regions.



Probability distribution generator

Grid Size 128 x 128

Type

Non-Uniform generator (LogNormal-Gaussian)
Uniform density generator

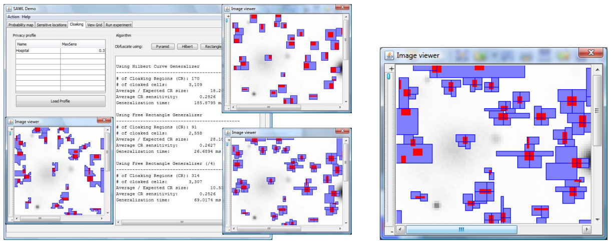
Import/export parameters

Parameter	Value
Random generator seed	34526329077870...
Population	1000000
Max number of attraction poi...	10000
Attraction point mu (mu1)	2
Attraction point sigma (sigma1)	2
Position std deviation mu (mu 2)	2
Attraction point sigma (sigma 2)	1

```
dico.probe.grid.generator.Log
params:
{AP_COUNT=10000, AP_MU=2, AP_
Seed: 3452632907787043840
Cell probability range:
[1.0919982300895695E-7,0.0021
```

x: [1.1E-7,0.0022]

Figure 2: a) Panel handling the Probability Grid (left); b) sensitive locations (right)



Privacy profile

Name	MaxSens
Hospital	0.3

Algorithm

Obfuscate using:

```
Using Hilbert Curve Generalizer
-----------------------------------------
# of Cloaking Regions (CR): 170
# of cloaked cells:         3,109
Average / Expected CR size:         18.2
Average CR sensitivity:        0.2826
Generalization time:         185.8795 m

Using Free Rectangle Generalizer
---------------------------------------------
# of Cloaking Regions (CR): 91
# of cloaked cells:         2,558
Average / Expected CR size:         28.1
Average CR sensitivity:        0.2627
Generalization time:         26.6894 ms

Using Free Rectangle Generalizer (/4)
-----------------------------------------
# of Cloaking Regions (CR): 314
# of cloaked cells:         3,307
Average / Expected CR size:         10.5
Average CR sensitivity:        0.2526
Generalization time:         69.0174 ms
```

Figure 3: a) Outcomes of the three methods (left); b) $Sens_{Div}$ with a more restrictive profile (right)

CRs generation phase. After completing the set-up, the user enters the privacy profile by specifying the sensitivity threshold for each sensitive location type (e.g., 0.3 for the hospital type) and then selects the cloaking method to run. The resulting CRs are displayed as blue polygons containing either entire or portions of sensitive locations while QoS measures are reported in a message window. The CRs obtained from different methods can be displayed in multiple Maps windows to allow a visual comparison of the results (Figure 3.a). In our example, the $Sens_{Div}$ algorithm generates smaller CRs than the other methods. To evaluate the method/s on a different scenario, for example with a more stringent privacy profile, it is sufficient to modify the profile and run again the cloaking methods. Figure 3.b shows the CRs generated by $Sens_{Div}$ with a more restrictive privacy profile. In summary, SAWL provides an extensible and flexible platform for location cloaking methods analysis. The software can be downloaded from the website: *http://www.silv.eu/SAWL*.

4. REFERENCES

[1] M.L. Damiani, E. Bertino, and C. Silvestri. The PROBE Framework for the Personalized Cloaking of Private Locations. *Transactions on Data Privacy*, (3)2:123–148, 2010.

[2] M. Musolesi and C. Mascolo. Designing mobility models based on social network theory. *SIGMOBILE Mobile Comput. Commun. Rev.*, 11(3):59–70, 2007.