

Trade-offs in location data accuracy and protection

Marco Cremonini

Dipartimento di Tecnologie dell'Informazione
Università degli Studi di Milano - Italy

First International Symposium on Global Information Governance 2008 - March 13-14, Pisa

Outline

Summary of two projects:

- 1 • Location-based Access Control System (LBAC)
 - Privacy-oriented language and policies
- 2 • Obfuscation-based location privacy
 - Tools and techniques

Cross-organizational issues arise integrating the two in a coherent location-based, privacy-aware framework for mobile applications

Outline

Summary of two projects:

- 1 • Location-based Access Control System (LBAC)
 - Privacy-oriented language and policies
- 2 • Obfuscation-based location privacy
 - Tools and techniques

Cross-organizational issues arise integrating the two in a coherent location-based, privacy-aware framework for mobile applications

Goals - LBAC

Provide a uniform privacy-aware access control system allowing parties to

- specify and enforce access regulations on information/services and privacy requirements over personally identifiable information (PII)
- specify and enforce restrictions on secondary use of private information after its release to external parties
- specify and enforce location-based access control policies

Motivation - Location Privacy

Privacy becomes a key aspect in computer security

- huge amount of personal data stored in digital form, with or without the consent of the owners
- huge data losses and mismanagement
- location information as one of the newest categories of personal information

Accuracy Vs. Privacy

Physical user position used in location-based service provisioning

- + it improves access control
- it affects user privacy

Need to balance:

- privacy needs (preferences) of users
- location accuracy needed by the services/applications

Open Issue

Provide a location privacy solution that protects privacy of the users still preserving the accuracy of users location information

Accuracy Vs. Privacy

Physical user position used in location-based service provisioning

- + it improves access control
- it affects user privacy

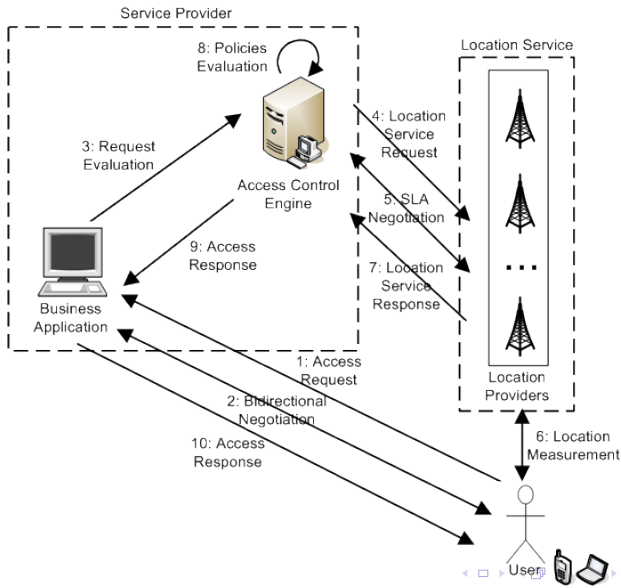
Need to balance:

- privacy needs (preferences) of users
- location accuracy needed by the services/applications

Open Issue

Provide a location privacy solution that protects privacy of the users still preserving the accuracy of users location information

Basic scenario



LBAC Requirements

- Depart from user authentication and provide interactive access control
- Maintain simplicity (logic languages powerful and expressive, but not usable by end users)
- Integrate seamlessly with emerging standards (e.g., SAML)
- Support semantics-aware conditions (based on defined ontologies)
- Allow for obligations and restrictions on secondary use

LBAC Privacy Policies

Different families of policies

- **access control policies** govern access to service and release of data stored at some service provider
- **release policies** govern release of personal private information
- **data handling policies** define restrictions on secondary use of PII
- **sanitized policies** regulate the dialog between parties to protect sensitive policy information

Access control and release language: Characteristics

- *Attribute-based restrictions*, policies definition based on properties of subjects and objects
- *XML-based syntax*, flexible and interoperable policies
- *Credential definition and integration*, requests for certified and uncertified data
- *Anonymous credentials support*, definition of conditions that can be satisfied by means of zero-knowledge proof
- *Support for context-based conditions*, definition of conditions based on physical position of the users and context information
- *Ontology integration*, generic assertions on subjects and objects

Location-based predicates

Type	Predicate	Description
Position	<code>inarea(<i>user</i>, <i>area</i>)</code>	Evaluate whether <i>user</i> is located within <i>area</i> .
	<code>disjoint(<i>user</i>, <i>area</i>)</code>	Evaluate whether <i>user</i> is located outside <i>area</i> .
	<code>distance(<i>user</i>, <i>entity</i>, <i>min_dist</i>, <i>max_dist</i>)</code>	Evaluate whether the distance between <i>user</i> and <i>entity</i> is within interval [<i>min_dist</i> , <i>max_dist</i>].
Movement	<code>velocity(<i>user</i>, <i>min_vel</i>, <i>max_vel</i>)</code>	Evaluate whether <i>user</i> 's speed falls within range [<i>min_vel</i> , <i>max_vel</i>].
Interaction	<code>density(<i>area</i>, <i>min_num</i>, <i>max_num</i>)</code>	Evaluate whether the number of users currently in <i>area</i> falls within interval [<i>min_num</i> , <i>max_num</i>].
	<code>local_density(<i>user</i>, <i>area</i>, <i>min_num</i>, <i>max_num</i>)</code>	Evaluate the density within a 'relative' area surrounding <i>user</i> .

Examples

- `inarea(Alice,Milan) = [True, 0.9, 2005-11-09-11:10am]`
- `velocity(Alice,70,90) = [True, 0.7, 2005-11-03-03:00pm]`
- `density(Director Office,0,1) = [False, 0.95, 2005-11-21-06:00pm]`

Relevance metric

Relevance $R \in (0, 1]$ is an adimensional, technology-independent metric of location information accuracy

- $\rightarrow 0$ when location information is considered unreliable
- $= 1$ when location information has best accuracy
- $\in (0, 1)$ when the location information accuracy is less than optimal for measurement errors and/or artificial degradations

Two relevance values characterize our LBAC solution

- evaluation relevance (\mathcal{R}_{Eval}), level of reliability of a location-based predicate evaluation
- LBAC relevance (\mathcal{R}_{LBAC}), minimum accuracy required for a location measurement or for a location-based predicate evaluation

A privacy-aware access control prototype

- Developed in the context of European project PRivacy and Identity Management for Europe (PRIME)
- Integrate access control, release, and data handling policies specification, evaluation, and enforcement

Location privacy - Working assumptions

- The area returned by a location measurement $Area(r_{meas}, x_C, y_C)$ is *planar* and *circular*
 - $Area(r_{meas}, x_C, y_C)$: circular area centered on the geographical coordinates (x_C, y_C) with radius r_{meas}
 - $Area(r_{meas}, x_C, y_C)$ includes user position with probability equal to 1
- The probability that the real user's position (x_U, y_U) belongs to a neighborhood of a random point (\hat{x}, \hat{y}) in $Area(r_{meas}, x_C, y_C)$ is uniformly distributed

Relevance

Estimates degrees of accuracy (R) and privacy ($1 - R$)

Location privacy solution based on three relevances

- initial relevance (R_{Init}): technological accuracy of a user location measurement as returned by a sensing technology

$$R_{Init} = \frac{r_{opt}^2}{r_{meas}^2}$$

- final relevance (R_{Final}): accuracy of the final obfuscated area produced by satisfying a user privacy preference

$$R_{Final} = \lambda R_{Init}$$

- intermediate relevance (R_{Inter}): relevance associated with the intermediate obfuscated area when a double obfuscation is used

Obfuscation techniques

Transform initial area R_{Init} into another area with relevance R_{Final}

Different techniques

- enlarging the radius (E)
- reducing the radius (R)
- shifting the center (S)
- combination of the above (double obfuscation)

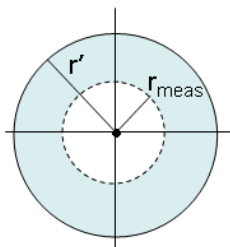
Obfuscation by enlarging the radius

Obfuscate a location measurement by enlarging the radius of the circular area (from r to $r' > r$)

- obfuscation derives from the decreasing of the probability density function (pdf) of the obfuscated area

$$R_{Final} = \frac{f_{r'}(x,y)}{f_r(x,y)} \cdot R_{Init} = \frac{r_{meas}^2}{r'^2} \cdot R_{Init}$$

$$\Rightarrow r' = r_{meas} \sqrt{\frac{R_{Init}}{R_{Final}}}$$

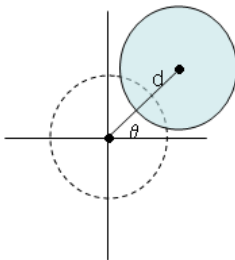


Obfuscation by shifting the center - 1

Obfuscate a location measurement by shifting the center of the area to a distance d

Decreased probability that the user belongs to the area

- distance $d \in [0, 2r_{meas}]$
- angle θ is randomly selected



Obfuscation by shifting the center - 2

Need to consider two probabilities:

- the probability that the real user's position belongs to the intersection between the original and obfuscated areas
- the probability that a random point selected from the whole obfuscated area belongs to the intersection

$$R_{Final} = \frac{Area_{Init \cap Final} \cdot Area_{Init \cap Final}}{Area(r_{meas}, x_c, y_c) \cdot Area(r_{meas}, x_c + \Delta x, y_c + \Delta y)} \cdot R_{Init}$$

$$\implies \frac{R_{Final}}{R_{Init}} = \frac{Area_{Init \cap Final}^2}{Area^2(r_{meas}, x_c, y_c)}$$

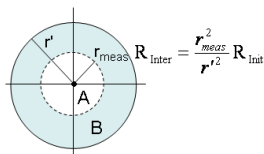
Given πr^2 as the value of both areas, $Area_{Init \cap Final} = \pi r^2 \frac{R_{Final}}{R_{Init}}$
and distance d can be calculated numerically

Double obfuscation

Combines different steps of enlarging radius, reducing radius, shifting center

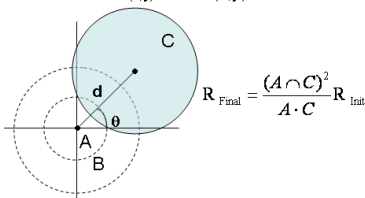
- any obfuscation can always be reduced to two steps
- obfuscated area should intersect with original area

1st obfuscation: $A(r) \xrightarrow{E} B(r')$



$$R_{\text{Inter}} = \frac{r_{\text{meas}}^2}{r'^2} R_{\text{Init}}$$

2nd obfuscation: $B(x,y) \xrightarrow{S} C(x',y')$



$$R_{\text{Final}} = \frac{(A \cap C)^2}{A \cdot C} R_{\text{Init}}$$

- Set of available techniques $\Sigma = \{E, R, S, SR, RS, ES, SE\}$
 - Σ is *minimum* and *complete*

Adversary model - 1

Relevance is not enough to measure the real privacy protection

Robustness of each obfuscation technique evaluated with respect to de-obfuscation attempts performed by adversaries

Issues to be considered in obfuscation robustness analysis:

- whether an adversary can manipulate an obfuscated area and obtain a more accurate location
- whether an adversary can evaluate the resulting accuracy gain or loss after de-obfuscation attempt

Adversary model - 2

Adversary knows the obfuscated location, the location sensing technology, and available obfuscation techniques

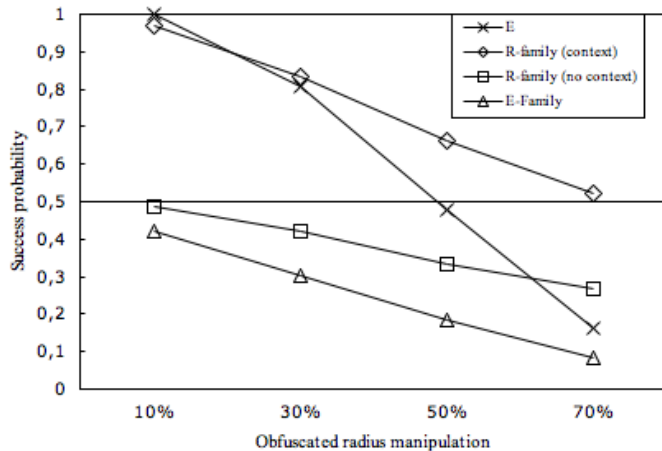
Based on contextual information adversary can infer the adoption of:

- R-family = $\{R, RS, SR\}$
- E-family = $\{E, ES, SE\}$

Adversary tries to reduce the obfuscation effects by:

- reducing the radius
- enlarging the radius

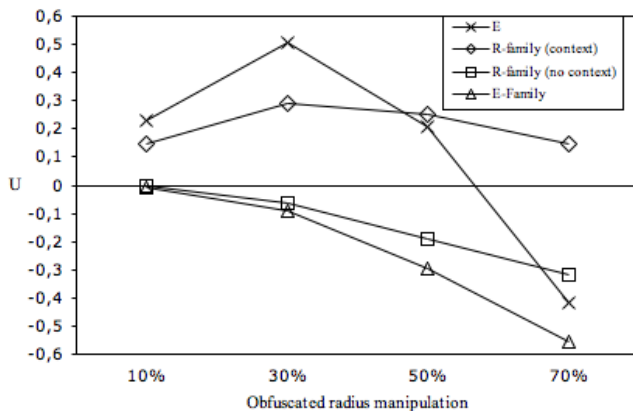
Success rate



Utility function

$$U = WG - (1 - W)L$$

where W is the rate of success, G is the mean gain, L is the mean loss.



To do

- Game theory approach
- More complex shapes and map constraints
- Path protection