tipsall
Xy-pic version 00 ¡00

Location Privacy Protection Through Obfuscation-based Techniques

C.A. Ardagna, M. Cremonini, E. Damiani,
S. De Capitani di Vimercati, and P. Samarati

Dipartimento di Tecnologie dell'Informazione
Università di Milano – 26013 Crema - Italy
{ardagna,cremonini,damiani,decapita,samarati}@dti.unimi.it

Abstract. The widespread adoption of mobile communication devices combined with technical improvements of location technologies are fostering the development of a new wave of applications that manage physical positions of individuals to offer location-based services for business, social or informational purposes. As an effect of such innovative services, however, privacy concerns are increasing, calling for more sophisticated solutions for providing users with different and manageable levels of privacy. In this work, we propose a way to express users privacy preferences on location information in a straightforward and intuitive way. Then, based on such location privacy preferences, we discuss a new solution, based on obfuscation techniques, which permits us to achieve, and quantitatively estimate through a metric, different degrees of location privacy.

1 Introduction

Information regarding physical locations of individuals is rapidly becoming easily available for processing by online and mobile *Location-Based Services* (LBSs). Customer-oriented applications, social networks and monitoring services can be functionally enriched with data reporting where people are, how they are moving or whether they are close by specific locations. To this end, several commercial and enterprise-oriented LBSs are already available and have gained popularity [4]. Key to those new LBSs are modern location technologies that have reached good precision and reliability at costs that most people (e.g., the cost of mobile devices) and companies (e.g., the cost of integrating location technologies in existing telecommunication infrastructures) can economically sustain.

Combined with novel application opportunities, however, threats to personal privacy are ramping up [4], as witnessed by recent security incidents targeting privacy of individuals, revealed faulty data management practices, and unauthorized trading of users personal information (including ID thefts and unauthorized profiling). Location information is not immune from such threats and presents new dangers such as stalking or physical harassment.

In this scenario, a novel contribution of the paper is represented by a comprehensive solution aimed at preserving location privacy of individuals through artificial perturbations of location information collected by sensing technology. In particular, location information of users is managed by a trusted middleware [5, 6, 9], which enforces users privacy through obfuscation-based techniques.

Key to this work is the concept of *relevance* as the adimensional metric for the location accuracy. A relevance value is always associated with locations and it quantitatively characterizes the degree of privacy artificially introduced into a location measurement. Based on relevance, it is possible to strike a balance between the need of service providers, requiring a certain level of location accuracy, and the need of users, asking to minimize the disclosure of personal location information. Both needs can be expressed as relevances and either quality of online services or location privacy can be adjusted, negotiated or specified as contractual terms.

The remainder of this paper is organized as follow. Section 2 presents related work. Section 3 discusses our working assumptions. Section 4 illustrates our approach for defining location privacy preferences and introduces the concept of *relevance*. Section 5 presents our obfuscation techniques. Section 6 describes a first solution to their composition and presents some examples of application. Section 7 gives our conclusions.

2 Related work

Location privacy issues are the subject of growing research efforts. The main branch of current research on LBS privacy focuses on users anonymity or partial identities [6, 7, 9]. Beresford and Stajano [6] present *mix zones*, a method used to enhance privacy in LBSs managed by trusted middlewares. The solution is based on preset physical zones where all users are indiscernible from one another. However, this solution is suitable for services that track users movement rather than, as in our case, for services requiring a user location at a specific time. Bettini et. al. [7] propose a framework in charge of evaluating the risk of sensitive location-based information dissemination, and a technique aimed at supporting k-anonymity [14]. Gruteser and Grunwald [9] define k-anonymity in the context of location obfuscation and propose a middleware architecture and an adaptive algorithm for adjusting location information resolution according to anonymity requirements.

Other works study the possibility of protecting users privacy through the definition of complex rule-based policies [10, 11]. Although policies-based solutions are suitable for privacy protection, users, often, are not willing to directly manage complex policies and, hence, refuse participation in pervasive environments. By contrast, in this work we have implemented a solution for expressing privacy preferences that is simple and intuitive.

Finally, the line of research closest to our work consists in the adoption of obfuscation techniques aimed at location privacy protection. Location obfuscation is complementary to anonymity. In particular, rather than anonymizing users identities, obfuscation-based solutions assume the identification of users and introduce perturbations into collected locations to decrease their accuracy. Duckham and Kulik [8] develop an obfuscation technique for protecting location privacy by artificially inserting into measurements some fake points with the same probability as the real user position. The paper proposes a formal framework providing a mechanism for balancing between user needs for high-quality information services and for location privacy. The work of Bellavista et al. [5] is based on points of interest with symbolic location granularity (e.g., city, country). This forces the privacy level to some predefined choices only, resulting in an excessively rigid solution.

Current obfuscation-based solutions have some shortcomings that our proposal tries to address. First, they do not provide a quantitative estimation of the actual privacy level, which makes them highly dependent on the application contexts and difficult to integrate into a full fledged location-based application scenario [1, 3]. Next, just a single obfuscation technique is usually implemented. By contrast, our work introduces the concept of relevance as an adimensional metric for location accuracy, defines more obfuscation techniques and demonstrate the benefits of their composition.

3 Working assumptions

Our work is based on two working assumptions that simplify our analysis with no loss of generality. Our first working assumption concerns the shape of a location measurement: *the area returned by a location measurement is planar and circular*. User location information, in fact, is affected by an intrinsic measurement error introduced by sensing technologies, resulting in spatial areas rather than geographical points. This assumption represents a particular case of the general requirement of considering convex areas and a good approximation for actual shapes resulting from many location technologies (e.g., cellular phones location). According to this assumption, a location measurement is defined as follows.

Definition 1 (Location measurement) *A location measurement of a user u is a circular area $Area(r, x_c, y_c)$, centered on the geographical coordinates (x_c, y_c) and with radius r, which includes the real user's position (x_u, y_u) with probability $P((x_u, y_u) \in Area(r, x_c, y_c)) = 1$.*

Definition 1 comes from observing that sensing technologies based on cellular phones usually guarantee that the real user's position falls within the returned area.

To discuss the effects of obfuscation techniques, we introduce our second assumption. Consider a random location within a location measurement $Area(r, x_c, y_c)$, where a "random location" is a neighborhood of random point $(\hat{x}, \hat{y}) \in Area(r, x_c, y_c)$. Our second assumption is that the probability that the real user's position (x_u, y_u) belongs to a neighborhood of a random point $(\hat{x}, \hat{y})$ is uniformly distributed over the whole location measurement. Accordingly, the joint probability density function (pdf) of the real user's position can be defined as follows.

Definition 2 (Uniform joint pdf) *Given a location measurement $Area(r, x_c, y_c)$, the joint probability density function (joint pdf) $f_r(x, y)$ of real user's position (x_u, y_u) to be in the neighborhood of point (x, y) is:*

$$f_r(x, y) = \begin{cases} \frac{1}{\pi r^2} & \text{if } x, y \in Area(r, x_c, y_c) \\ 0 & \text{otherwise.} \end{cases}$$

Same assumption can be found in other works on this topic [12]. In this work, assuming a uniform distribution simplifies the discussion with no loss of generality. Considering Gaussian-like distributions, the consequence on our work is that obfuscating simply by scaling the radius of a location measurement is ineffective, while stronger obfuscation effects can be still achieved by combining different techniques.

4 Privacy preferences and location relevance

The ultimate goal of this work is to design a solution able to manage location privacy as a functional term, required or adjusted by users according to their preferences and application context, and negotiated as a service attribute by users and LBSs. To this end, location privacy should be measured and quantified with regard to the *accuracy* of a user position, that is, the more accurate the position, the less privacy. Furthermore, location privacy should be measured regardless of specific application contexts and should be expressed quantitatively as a service parameter without sticking to some coarse-grained preset meta-locations such as "city" or "department", which represents simplified instances of privacy preferences that should be supported by a most general and flexible solution.

Before defining obfuscation techniques and their combination, we discuss some key aspects: accuracy estimations of available location technologies, the specification of users privacy preferences, and the concept of relevance.

4.1 Location accuracy and measurement quality

The accuracy of a location measurement necessarily depends on the specific sensing technology and on the environmental conditions. Several works describe available sensing technologies discussing their accuracy. In [15], the authors provide a survey of standard positioning solutions for cellular networks such as, *E-OTD* for GSM, *OTDOA* for Wideband CDMA (WCDMA), and *Cell-ID*. Specifically, E-OTD location method is based on the existing observed time difference (OTD) feature of GSM systems. The accuracy of the E-OTD estimation, in recent studies, has been found to range from 50m to 125m. Observed Time Difference Of Arrival (OTDOA), instead, is designed to operate over wideband-code division multiple access (WCDMA) networks. The positioning process achieves a location accuracy of 50m at most. Finally, Cell-ID is a simple positioning method based on cell sector information, where cell size varies from 1-3km in urban areas to 3-20km in suburban/rural areas.

To evaluate the quality of a given location measurement, its accuracy must be compared with the nominal accuracy that the adopted sensing technology can reach. To this end, we call r_{meas} the radius of a measured area and r_{opt} the radius of the area that would be produced if the best accuracy is reached. In other words, r_{meas} represents the *actual measurement error*, while r_{opt} is the *minimum error*. Therefore, the ratio r_{opt}^2/r_{meas}^2 is a good estimation of the quality of each location measurement. For instance, assume that a user position is located with accuracy r_{meas}=62.5m using E-OTD method, accuracy r_{meas}=50m using OTDOA, and accuracy r_{meas}=1km using Cell-ID. Optimal accuracy is r_{opt}=50m. Then, according to r_{opt}^2/r_{meas}^2, the area provided by OTDOA has a measurement quality of 1, whereas the others have a quality proportionally reduced to 0.8 for E-OTD, and 0.05 for Cell-ID.

4.2 User privacy preferences

Systems that want to let users express their privacy preferences must strike a balance between the two traditionally conflicting requirements of usability and expressiveness. Complex policy specifications, fine-grained configurations and explicit technological details discourage users from fully exploiting the provided functionalities. Our goal is then to allow users to express privacy preferences in an intuitive and straightforward way. Our solution is based on users privacy preferences specified as a *minimum distance* [8, 13]. According to this setting, for example, a user can define "100 meters" as her privacy preference, which means that she demands to be located with an accuracy not better than 100 meters. Considering circular areas, the privacy requirement is implemented by enlarging the radius of the original measurement to 100 meters, at least. However, privacy preferences expressed as a minimum distance have the drawback of being meaningful if associated with a technique that enlarge the original measurement only. Another issue that is often neglected by traditional location obfuscation solutions is the possibility to compose different obfuscation techniques to increase their robustness with respect to possible de-obfuscation attempts performed by adversaries.

Therefore, a major challenge is to design a system able to integrate several obfuscation techniques still relying on the definition of privacy preference in its simplest form, e.g., it would be unrealistic to explicitly ask user to specify a particular composition of techniques. Our solution transforms a simple preference like a minimum distance into a more general functional term with the constraint that the final obfuscated area produced by different obfuscation techniques must be equivalent, in terms of location privacy, to the area that would be derived by just enlarging the radius of the original measurement to the specified minimum distance. This way, we let users specify their preferences in the most intuitive way, whereas we can adopt obfuscation techniques different and more robust than the simple radius enlargement. The availability of a single obfuscation technique by enlarging the radius, in fact, gives to an adversary the possibility of guessing a better user position by simply reducing the observed area. Our solution, instead, introduces additional obfuscation techniques, and therefore improves user privacy.

To this end, we first introduce the attribute λ that represents a *relative privacy preference* (or, in other terms, a relative degradation of the location accuracy). λ must be derived from the minimum distance specified by a user, which we call r_{min}, and from the radius of the original measurement, the previously introduced r_{meas}. Having assumed circular areas, the relative accuracy degradation obtained by setting r_{min} is:

$$\lambda = \frac{max(r_{meas}, r_{min})^2 - r_{meas}^2}{r_{meas}^2} = \frac{max(r_{meas}, r_{min})^2}{r_{meas}^2} - 1 \tag{1}$$

The term $max(r_{meas}, r_{min})$ represents the special case of a minimum distance r_{min} smaller than the original r_{meas}. This is possible because the user is not aware of the actual accuracy of sensing technologies and the original measure could already satisfy the privacy preference by itself. Accordingly, the term λ is zero when the measurement accuracy (i.e., r_{meas}) already satisfies the user requirement (i.e., r_{min}) and no transformation to the original measurement is needed to satisfy privacy preferences. Otherwise, when this is not the case (i.e., $r_{min} > r_{meas}$), λ corresponds to various degrees of accuracy degradation, e.g., $\lambda = 0.2$ means 20% of degradation, $\lambda = 1$ means 100% of degradation and any value $\lambda > 1$ corresponds to a degradation greater than 100%.

Up to this point, the first benefit achieved by deriving λ from r_{min} and r_{meas} is that we can process a privacy preference as a relative degradation rather than the strictly dimensional and tightly coupled with the enlargement of the measured area r_{min}.

The next step is to introduce other obfuscation techniques and select them, individually or combined, to produce an obfuscated area that degrades the original accuracy as imposed by λ. This way, we can employ an enriched set of obfuscation techniques still relying on the simple definition of r_{min} as the user privacy preference. The drawback, which we consider acceptable, is that we are changing the meaning of the user preference r_{min}, which is not necessarily the radius of the obfuscated area. Instead, it represents a logical constraints that can be informally expressed as: *the location area produced by one or more a priori undetermined obfuscation techniques must be equivalent, in term of privacy, to the one produced by enlarging the radius of the original measurement up to r_{min}.*

4.3 Relevance

Key to our work is the notion of *relevance*, defined as an adimensional, technology-independent metric for the accuracy of an obfuscated area. The relevance metric is a value $\mathcal{R} \in (0, 1]$ that tends to 0 when location information must be considered unreliable for application providers; it is equal to 1 when location information has best accuracy; and a relevance value in (0,1) corresponds to some degrees of accuracy. Accordingly, the *location privacy* provided by an obfuscated location is evaluated by (1-$\mathcal{R}$). The reason for choosing to represent the accuracy of a location as a primitive concept rather than the privacy is functional. We assume that LBSs have to manage locations that, on the one side, could be perturbed for privacy reasons, while on the other side could be required to have an accuracy not below a threshold to preserve a certain quality of service. In our solution, all locations have an associated relevance attribute, from an initial location affected by a measurement error of sensing technologies to all possible subsequent manipulations to provide privacy. This way, relevance is the general functional term that qualifies a location with respect to either accuracy or privacy requirements. Two important relevance values characterize our privacy management solution:

- *Initial relevance ($\mathcal{R}_{Init}$).* The metric for the accuracy of a user location measurement as returned by a sensing technology. This is the initial value of the relevance that only depends on the intrinsic measurement error.
- *Final relevance ($\mathcal{R}_{Final}$).* The metric for the accuracy of the final obfuscated area produced by satisfying a relative privacy preference λ. It is derived, starting by the initial relevance, through the application of one or more obfuscation techniques.

A third relevance value is used when the combination of techniques will be discussed. It represents the *intermediate relevance*, denoted $\mathcal{R}_{Inter}$, derived by applying the first of two obfuscation techniques.

With regard to $\mathcal{R}_{Init}$, it evaluates the accuracy of the actual area returned by a specific location measurement. A good metric is the ratio of the area that would have been returned if the best accuracy was achieved (i.e., the one with radius r_{opt}) and the actual measured area (i.e., the one with radius r_{meas}). $\mathcal{R}_{Final}$ instead, is derived from $\mathcal{R}_{Init}$ by considering the relative privacy preference λ.

Definition 3 ($\mathcal{R}_{Init}$ and $\mathcal{R}_{Final}$) *Given a location measurement area of radius r_{meas} measured by a sensing technology, a radius r_{opt} representing the best accuracy of sensing technologies and a relative privacy preference λ, initial relevance $\mathcal{R}_{Init}$ and final relevance $\mathcal{R}_{Final}$ are calculated as:*

$$\mathcal{R}_{Init} = \frac{r_{opt}^2}{r_{meas}^2} \tag{2}$$

$$\mathcal{R}_{Final} = \frac{\mathcal{R}_{Init}}{\lambda + 1} \tag{3}$$

These definitions represent, respectively, our general forms of $\mathcal{R}_{Init}$ and $\mathcal{R}_{Final}$. By definition of λ (see (1)), the term $\frac{1}{\lambda+1}$ represents the degradation of the initial $\mathcal{R}_{Init}$ that satisfies the user privacy preference. The corresponding obfuscated area will be qualified by relevance $\mathcal{R}_{Final}$. In equation (3), substituting the term $\mathcal{R}_{Init}$ with equation (2) and term λ with equation (1), it results that $\mathcal{R}_{Final} = \frac{r_{opt}^2}{r_{min}^2}$, assuming $r_{min} > r_{meas}$ in (1). This represents the value of $\mathcal{R}_{Final}$ that corresponds to degrading the accuracy by λ, as for user's privacy preference.

5 Obfuscation techniques

We now present three basic obfuscation techniques and their operators. Since there could be one or two obfuscation steps in our solution, we generically call $\mathcal{R}$ the relevance associated with the area to be obfuscated and $\mathcal{R}'$ the relevance of the obfuscated area. If only one obfuscation step is performed, then $\mathcal{R}$ =$\mathcal{R}_{Init}$ and $\mathcal{R}'$ =$\mathcal{R}_{Final}$. For two obfuscation steps, we have $\mathcal{R}$ =$\mathcal{R}_{Init}$ and $\mathcal{R}'$ =$\mathcal{R}_{Inter}$ for the first one, and $\mathcal{R}$ =$\mathcal{R}_{Inter}$ and $\mathcal{R}'$ =$\mathcal{R}_{Final}$ for the second one.

Furthermore, we employ *obfuscation operators* as a logical representation of the physical transformations realized by different obfuscation techniques: *i)* the `Enlarge` operator (E) degrades the accuracy of an initial location area by enlarging its radius; *ii)* the `Shift` operator (S) degrades the accuracy of an initial location area by shifting its center; and *iii)* the `Reduce` operator (R) degrades the accuracy of an initial location area by reducing its radius.

5.1 Obfuscation by enlarging the radius

Obfuscating a location measurement area by increasing its radius (see Fig. 1(a)) is the technique that most current solutions adopt. Obfuscation is a probabilistic effect provided by the decreasing of the joint probability density function (pdf), which we can express as $\forall r, r' \in \mathbb{R}^+, r < r' : f_r(x,y) > f_{r'}(x,y)$. The relevance $\mathcal{R}'$ can be derived from $\mathcal{R}$ by using the ratio of the associated pdf as the scalar factor:

$$\mathcal{R}' = \frac{f_{r'}(x,y)}{f_r(x,y)} \cdot \mathcal{R} = \frac{r^2}{r'^2} \cdot \mathcal{R}, \quad \textit{with } r < r' \tag{4}$$

Therefore, given two relevances, $\mathcal{R}$ and $\mathcal{R}'$, and the radius r of the initial area, an obfuscated area calculated with this technique has a final radius: $r' = r\sqrt{\frac{\mathcal{R}}{\mathcal{R}'}}$.

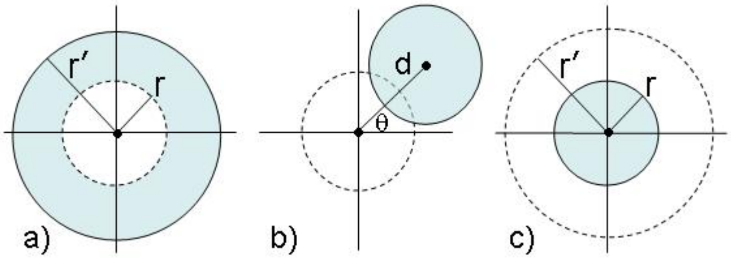


Fig. 1. Obfuscation by enlarging the radius (a), shifting the center (b), and reducing the radius (c)

Finally, radius r' can be expressed as a function of λ: $r' = r\sqrt{\lambda + 1}$. This result is straightforward from equations (3) and (4) and reflects the definition of λ, which depends from r_{min} and assumes an obfuscation by enlarging the radius.

For instance, let the user privacy preference be r_{min}=1 km. Suppose that the location measurement of a user u has radius r_{meas}=0.5 km, and the optimal measurement accuracy is r_{opt}=0.4 km. Given this information, relevance $\mathcal{R}_{Init}$ associated with the location measurement, and relative privacy preference λ are calculated as $\mathcal{R}_{Init}=\frac{r^2_{opt}}{r^2_{meas}}=0.64$, and $\lambda = \frac{\max(r_{meas},r_{min})^2}{r^2_{meas}} - 1=3$, respectively. Having calculated the relative privacy preference λ, $\mathcal{R}_{Final}$ is derived as $\mathcal{R}_{Final}=(\lambda + 1)^{-1}\mathcal{R}_{Init}=0.16$. Now, the obfuscation by enlarging the radius is applied and the obfuscated area is derived by calculating $r' = r_{meas}\sqrt{\lambda + 1}=1$ km. Note that, since a single obfuscation by enlarging the radius is used, $r' = r_{min}$. However, this example shows the computations that have to be applied when a double obfuscation is used (see Section 6).

5.2 Obfuscation by shifting the center

Shifting the center of a location measurement is another viable obfuscation technique (see Fig. 1(b)). An obfuscated area is derived from the original area by calculating the distance d between the two centers [2]. To measure the obfuscation effect and define the relation between relevances, two probabilities must be composed: *i)* the probability that the real user's position belongs to the intersection $Area_{Init \cap Final}$, and *ii)* the probability that a random point selected from the whole obfuscated area belongs to the intersection. Then, the relation between relevances $\mathcal{R}$ and $\mathcal{R}'$ is represented by:

$$\mathcal{R}' = P((x_u, y_u) \in Area_{Init \cap Final}) \cdot P((x, y) \in Area_{Init \cap Final}) = \\ \frac{Area_{Init \cap Final}}{Area(r, x_c, y_c)} \cdot \frac{Area_{Init \cap Final}}{Area(r, x_c + \Delta x, y_c + \Delta y)} = \frac{Area^2_{Init \cap Final}}{Area(r, x_c, y_c)^2} \cdot \mathcal{R} \qquad (5)$$

Recalling equations (3) and (5), it follows that $(\lambda + 1)^{-1} = \frac{Area^2_{Init \cap Final}}{Area(r, x_c, y_c)^2}$. Then, given λ, and πr^2 as the value of both areas, the overlapping can be expressed as: $Area_{Init \cap Final} = \pi r^2 / \sqrt{\lambda + 1}$.

Distance d between the centers is the unknown variable to be derived to obtain the obfuscated area. It can be calculated by expanding the term $Area_{Init \cap Final}$ as a function of d and by solving the following system of equations, whose variables are d, σ and γ. Here, σ and γ are the central angles of circular sectors identified by the two radii connecting the centers of the areas with the intersection points of original and obfuscated areas.[1]

$$\begin{cases} \left[\frac{\sigma}{2} r^2 - \frac{r^2}{2} \sin \sigma\right] + \left[\frac{\gamma}{2} R^2 - \frac{R^2}{2} \sin \gamma\right] = \sqrt{\delta} \pi r \cdot R \\ d = r \cos \frac{\sigma}{2} + R \cos \frac{\gamma}{2} \\ r \sin \frac{\sigma}{2} = R \sin \frac{\gamma}{2} \end{cases} \qquad (6)$$

Solutions of this system can be obtained numerically. By our definitions, obfuscated areas calculated by shifting the center satisfy a relative privacy preference λ and thus provides same privacy of an obfuscated area that would have been calculated with an enlarged radius r_{min}.

For instance, let the user specifies her privacy preference through r_{min}=1.42 km. Suppose that the location measurement of a user u has radius r_{meas}=1 km, and the optimal measurement accuracy is r_{opt}=0.8 km. Relevance $\mathcal{R}_{Init}$ and λ are calculated as $\mathcal{R}_{Init}$=0.64, and λ=1, respectively. Then, $\mathcal{R}_{Final}$=0.32 is derived and the obfuscation by shifting the center applied. At this point, distance d=0.464 km is calculated by solving the system of equation (6). Finally, an angle θ is randomly selected and the obfuscated area is generated.

5.3 Obfuscation by reducing the radius

The third obfuscation technique consists in reducing the radius r of one location from r to r', as showed in Fig. 1(c). The obfuscation effect is

[1] The system of equation (6) is presented in the most general form, where there are two areas with different radii (i.e., r and R).

produced by a correspondent reduction of the probability to find the real user location within the returned area, whereas the joint pdf is fixed.

If we call (x_u, y_u) the unknown real user position coordinates, by assumption the probability that the real user position falls in the area of radius r is $P((x_u, y_u) \in Area(r, x, y)) = 1$. When we obfuscate by reducing the radius, an area of radius $r' \leq r$ is returned, where $P((x_u, y_u) \in Area(r', x, y)) \leq P((x_u, y_u) \in Area(r, x, y))$, since a circular ring having pdf greater than zero has been excluded.

With regard to relevances $\mathcal{R}$ and $\mathcal{R}'$, their relation can be defined as:

$$\mathcal{R}' = \frac{P((x_u, y_u) \in Area(r', x, y))}{P((x_u, y_u) \in Area(r, x, y))} \cdot \mathcal{R} = \frac{r'^2}{r^2} \cdot \mathcal{R}, \quad \textit{with } r' < r \tag{7}$$

From (3) and (7), it follows that $(\lambda + 1)^{-1} = \frac{r'^2}{r^2}$. Then, given λ and r, the area returned when obfuscation by reducing the radius is applied has radius: $r' = \frac{r}{\sqrt{\lambda+1}}$. Again, similarly to the previous technique, obfuscated areas calculated in this way by reducing the radius satisfy, according to our semantics, a relative privacy preference λ and consequently, also the corresponding user privacy preference r_{min}.

For instance, let the user privacy preference be r_{min}=1 km. Suppose that the location measurement of a user u has radius r_{meas}=0.5 km, and the optimal measurement accuracy is r_{opt}=0.4 km. Relevance $\mathcal{R}_{Init}$ and λ are calculated as $\mathcal{R}_{Init}$=0.64, and λ=3. $\mathcal{R}_{Final}$ is derived as $\mathcal{R}_{Final}=(\lambda + 1)^{-1}\mathcal{R}_{Init}$=0.16. Now, the obfuscation by reducing the radius is applied and the obfuscated area, respecting user privacy preference r_{min}, is derived by calculating $r' = \frac{r_{meas}}{\sqrt{\lambda+1}}$=0.25 km.

6 Double obfuscation

Given the obfuscation techniques just introduced, users privacy preferences can be satisfied either by using one technique among the three or by composing two techniques. In the last case, an obfuscated area produced by one obfuscation technique (operator h) is further obfuscated by the application of a second technique (operator g). Formally, let be $\mathcal{A}$ the set of location areas, $h : \mathcal{A} \rightarrow \mathcal{A}$ and $g : \mathcal{A} \rightarrow \mathcal{A}$ be two obfuscation operators, where the areas produced by applying the operator h are the inputs of the second operator g, which finally produces the obfuscated areas. Recalling that the ultimate goal of an obfuscation process is to reduce the location accuracy estimated by an initial relevance $\mathcal{R}_{Init}$ to a final relevance $\mathcal{R}_{Final}$, in case of double obfuscation, the intermediate

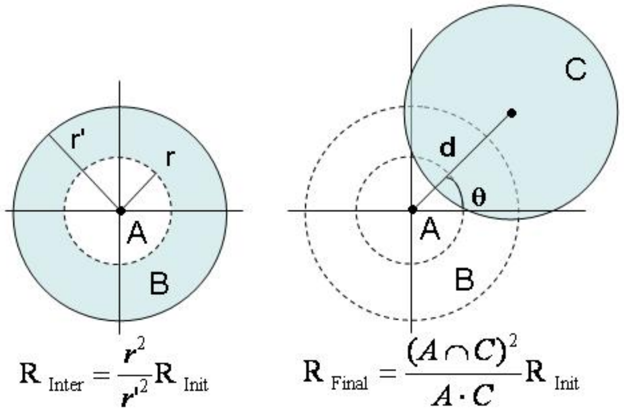
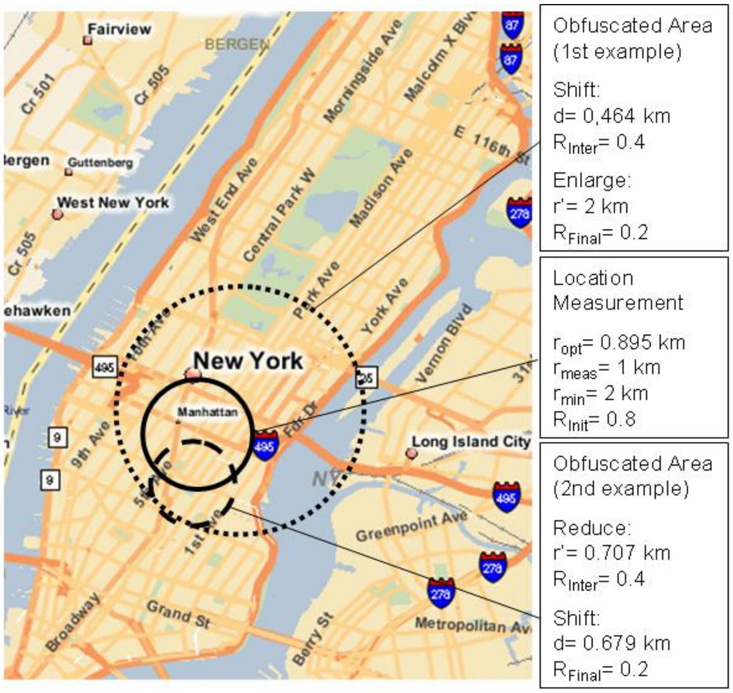

r'
r
A
B
C
d
θ
A
B
$R_{\text{Inter}} = \frac{r^2}{r'^2} R_{\text{Init}}$
$R_{\text{Final}} = \frac{(A \cap C)^2}{A \cdot C} R_{\text{Init}}$

Fig. 2. Example of $E - S$ obfuscation

term $\mathcal{R}_{Inter}$ must be introduce to represent the relevance achieved by the first obfuscation step.

As a special example of obfuscation composition, let us consider an enlargement of the radius followed by a shift of the center (see Fig. 2). For the first obfuscation, the radius of area B is calculated by applying operator `Enlarge` and equation $\mathcal{R}' = \frac{r^2}{r'^2} \cdot \mathcal{R}$, with $\mathcal{R} = \mathcal{R}_{Init}$ and $\mathcal{R}' = \mathcal{R}_{Inter}$. For the second step, operator `Shift` is used with an important constraint: the domain of the `Shift` operator must be restricted to those areas that have an intersection with the *original* measured area (i.e., in our example the intersection between the final obfuscated area C and area A should not be empty). As a consequence, to calculate the final obfuscated area C, we need to determinate distance d, which depends on the overlap between area A and C. The reason is that to respect privacy preference λ, the second operator of a composition must be always referred to the original measured area A. Accordingly, equation $\mathcal{R}' = \frac{(A \cap C)^2}{A \cdot C} \cdot \mathcal{R}$, has $\mathcal{R}' = \mathcal{R}_{Final}$ and $\mathcal{R} = \mathcal{R}_{Init}$, rather than $\mathcal{R} = \mathcal{R}_{Inter}$ as we would have expected in general.

Finally, we observe that, whereas in theory it is possible to compose operators E, R, and S in an indeterminate number of steps, there is never any convenience to combine more than two techniques. This follows by a geometric property of circles assuring that, given two circles A_1 and A_2, A_2 can be generated starting from A_1 through two geometric operations at most: one center-shifting, and one between radius enlargement or reduction. Finally, we observe that, as for most composable functions, the *commutative property* does not hold for the composition of opera-



Obfuscated Area (1st example)

Shift:
d= 0,464 km
R_{Inter}= 0.4

Enlarge:
r'= 2 km
R_{Final}= 0.2

Location Measurement

r_{opt}= 0.895 km
r_{meas}= 1 km
r_{min}= 2 km
R_{Init}= 0.8

Obfuscated Area (2nd example)

Reduce:
r'= 0.707 km
R_{Inter}= 0.4

Shift:
d= 0.679 km
R_{Final}= 0.2

Fig. 3. Example of $S - E$ and $S - R$ obfuscation on a large scale

tors `Enlarge` or `Reduce` with `Shift`. Therefore, the available obfuscation choices are: *i)* traditional single obfuscations E, R, and S; *ii)* double obfuscations $E - S$, $S - E$, $R - S$, and $S - R$.

6.1 Double obfuscation examples

We describe two examples of possible application of double obfuscation. For sake of clarity, we suppose an user located in Manhattan, around the *Empire State Building.* A location measurement is assumed to have radius r_{meas}=1km (see the area with filled line in Fig. 3).[2] The user has specified her privacy preference as r_{min}=2km. Given these information, λ, $\mathcal{R}_{Init}$, and $\mathcal{R}_{Final}$ are calculated before applying obfuscations.

For the first example, an $S - E$ obfuscation has been applied. The obfuscation process starts by setting $\theta = \pi/4$ and $\mathcal{R}_{Inter}$=0.4.

Distance d=0.464km is calculated by solving the system of equations (6), and location measurement area is shifted accordingly generating an

[2] In these examples, r_{opt}=0.895km. We are aware that this assumption is far from reality, but it was assumed for simplicity.

obfuscated area of relevance $\mathcal{R}_{Inter}$. Finally, the `Enlarge` operator is applied to the area with relevance $\mathcal{R}_{Inter}$, and final radius r'=2 km is computed to achieve relevance $\mathcal{R}_{Final}$. This way, when the user location is released to a LBS she results positioned in a bigger area that includes nearly all Central Manhattan (see the area with dotted line in Fig. 3).

For the second example, suppose a $R-S$ obfuscation. Again, for simplicity we set $\mathcal{R}_{Inter}$=0.4 and $\theta = 5\pi/3$. Radius r'=0.707 km is computed from (7) and the first obfuscated area is produced. Then, the system of equations (6) is solved numerically resulting in d =0.679 km, and the center is shifted. This way, when the user location is released to a LBS, the user seems located just around *Madison Square* (see the area with dashed line showed in Fig. 3).

7 Conclusions and Future Work

We presented privacy-enhanced techniques that protect user privacy based on spatial obfuscation. Our proposal aims at achieving a solution that both considers the accuracy of location measurements, which is an important feature of location information, and the need of privacy of users. In addition to several obfuscation techniques for privacy preservation, we also present and define a formal and intuitive way to express users privacy preferences, and a formal metric for location accuracy. Issues to be investigated include the analysis of our solution assuming Gaussian-like distributions, the evaluation of obfuscation techniques robustness against de-obfuscation attacks, and the possibility to manage different privacy preferences expressed by users.

Acknowledgments

This work was partially supported by the European Union within the PRIME Project in the FP6/IST Programme under contract IST-2002-507591, by the Italian Ministry of Research Fund for Basic Research (FIRB) under project RBNE05FKZ2 and by the Italian MIUR under project MAPS. We would like also to thank the anonymous referees for their helpful comments on the paper.

References

1. C.A. Ardagna, M. Cremonini, E. Damiani, S. De Capitani di Vimercati, and P. Samarati. Supporting location-based conditions in access control policies. In *Proc. of the ACM Symposium on Information, Computer and Communications Security (ASIACCS'06)*, Taipei, Taiwan, March 2006.

2. C.A. Ardagna, M. Cremonini, E. Damiani, S. De Capitani di Vimercati, and P. Samarati. Managing privacy in LBAC systems. In *Proc. of the Second IEEE International Symposium on Pervasive Computing and Ad Hoc Communications (PCAC-07)*, Niagara Falls, Canada, May 2007.
3. V. Atluri and H. Shin. Efficient enforcement of security policies based on tracking of mobile users. In *Proc. of the 20th Annual IFIP WG 11.3 Working Conference on Data and Applications Security*, pages 237–251, Sophia Antipolis, France, 2006.
4. L. Barkhuus and A. Dey. Location-based services for mobile telephony: a study of user's privacy concerns. In *Proc. of the 9th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2003)*, pages 709–712, Zurich, Switzerland, September 2003.
5. P. Bellavista, A. Corradi, and C. Giannelli. Efficiently managing location information with privacy requirements in wi-fi networks: a middleware approach. In *Proc. of the International Symposium on Wireless Communication Systems (ISWCS'05)*, pages 1–8, Siena, Italy, September 2005.
6. A. R. Beresford and F. Stajano. Mix zones: User privacy in location-aware services. In *Proc. of the 2nd IEEE Annual Conference on Pervasive Computing and Communications Workshops (PERCOMW04)*, 2004.
7. C. Bettini, X.S. Wang, and S. Jajodia. Protecting privacy against location-based personal identification. In *Proc. of the 2nd VLDB Workshop on Secure Data Management*, LNCS 3674, Springer-Verlag, 2005.
8. M. Duckham and L. Kulik. A formal model of obfuscation and negotiation for location privacy. In *Proc. of the 3rd International Conference PERVASIVE 2005*, Munich, Germany, May 2005.
9. M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proc. of the 1st International Conference on Mobile Systems, Applications, and Services*, 2003.
10. C. Hauser and M. Kabatnik. Towards Privacy Support in a Global Location Service. In *Proc. of the IFIP Workshop on IP and ATM Traffic Management (WATM/EUNICE 2001)*, Paris, France, 2001.
11. M. Langheinrich. A privacy awareness system for ubiquitous computing environments. In *Proc. of the 4th International Conference on Ubiquitous Computing (Ubicomp 2002)*, pages 237–245, September 2002.
12. M.F. Mokbel, C-Y. Chow, and W.G. Aref. The new casper: Query processing for location services without compromising privacy. In *Proc. of the 32nd International Conference on Very Large Data Bases*, pages 763–774, Korea, 2006.
13. Openwave. *Openwave Location Manager*, 2006. http://www.openwave.com/.
14. P. Samarati. Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):1010–1027, 2001.
15. G. Sun, J. Chen, W. Guo, and K.J. Ray Liu. Signal processing techniques in network-aided positioning: A survey of state-of-the-art positioning designs. *IEEE Signal Processing Magazine*, pages 12–23, July 2005.