

An Obfuscation-based Approach for Protecting Location Privacy

Claudio A. Ardagna, Marco Cremonini, Sabrina De Capitani di Vimercati, Pierangela Samarati

Abstract—The pervasive diffusion of mobile communication devices and the technical improvements of location techniques are fostering the development of new applications that use the physical position of users to offer location-based services for business, social, or informational purposes. In such a context, privacy concerns are increasing and call for sophisticated solutions able to guarantee different levels of location privacy to the users. In this paper, we address this problem and present a solution based on different *obfuscation operators* that, when used individually or in combination, protect the privacy of the location information of users. We also introduce an adversary model and provide an analysis of the proposed obfuscation operators to evaluate their robustness against adversaries aiming to reverse the obfuscation effects to retrieve a location that better approximates the location of the users. Finally, we present some experimental results that validate our solution.

Index Terms—Privacy, Obfuscation techniques, Location-based services



1 INTRODUCTION

The physical location of users is rapidly becoming easily available as a class of personal information that can be processed for providing new online and mobile services, generally called *Location-Based Services* (LBSs). Customer-oriented applications, social networks, and monitoring services can be greatly enriched with data reporting where people are, how they are moving, or whether they are close to specific locations. Several commercial and enterprise-oriented SBSs are already available and have gained popularity (e.g., [4], [13], [26]), driven by the relevant enhancements achieved in the field of sensing technologies. Location techniques permit to gather location information with good precision and reliability at costs that most people (e.g., the cost of current mobile devices like cellular phones) and companies (e.g., the cost of integrating location techniques in current telecommunication systems) can economically sustain.

In this context, the privacy of the users, which is already the center of many concerns for the risks posed by current online services [4], [29], [34], can be threatened by SBSs. The publicity gained by recent security incidents that have targeted the privacy of users has revealed faulty data management practices and unauthorized trading of personal information (including ID thefts and unauthorized profiling). For instance, legal cases have been reported, where rental companies used the GPS technology to track their cars and charge users for agreement infringements [9], or where an organization used a location service to track its own employees [25]. In addition, research on privacy issues has gained a

relevant boost since providers of online and mobile services have often largely exceeded in collecting personal information in the name of service provision.

In such a worrisome scenario, the concept of *location privacy* can be defined as *the right of individuals to decide how, when, and for which purposes their location information can be released to other parties*. The improper exposure of location information could result in severe consequences that make users the target of fraudulent attacks [15].

Current research on location privacy has mainly focused on supporting anonymity and partial identities [7], [8], [16], [19], [31]. To a certain extent, anonymity and complete knowledge of personal information are the opposite endpoints of all the degrees of personal information knowledge managed by online services, and location information is just one type of personal information that often needs to be bound to a user identity. Anonymity is however not viable in the provision of an online service when the identification of users is required [23]. In this case, a solution to protect the privacy of users consists in decreasing the accuracy of location information [14], [30]. As a matter of fact, many SBSs do not need to have available location information as accurate as possible to offer an acceptable quality of service to users.

In this paper, we present a novel solution aimed at preserving the location privacy of the users by perturbing location information measured by sensing technologies. We focus on the development of techniques for protecting a single sample of location information. For the sake of concreteness, we consider locations gathered by means of cellular phones as our reference, even if our solution is not bound to a specific location technique. One important characteristic of cellular phones is their large availability and the possibility to be used as a source of location information both indoor and outdoor

• The authors are with Dipartimento di Tecnologie dell'Informazione, Università degli Studi di Milano, 26013 Crema, Italy.
E-mail: firstname.lastname@unimi.it

(on the contrary, GPS is operating mainly outdoor). Key aspects of our perturbation process, called *obfuscation*, are *i*) to allow users to express their privacy preferences in a simple and intuitive way, and *ii*) to enforce the privacy preferences through a set of techniques robust against a relevant class of de-obfuscation attacks. To this end, we introduce the concept of *relevance* as a metric of both location information accuracy and privacy that abstracts from the physical attributes of the sensing technology as well as from the actual technique employed to obfuscate a location. This way, while users have just to select a relevance value, the robustness of the solution is guaranteed by randomly selecting one of the techniques to produce the obfuscated location. The robustness is demonstrated by our experiments simulating an attacker aiming at reversing the protection granted by obfuscation. Another benefit that the relevance metric could bring to LBSs is to support automated negotiation protocols handling the trade-off between the level of location accuracy for LBS provision requested by service providers and the protection of the location information requested by users. Both needs could be expressed as relevance and the quality of online services or the location privacy can be adjusted, negotiated, or specified as contractual terms to meet a certain relevance.

The remainder of this paper is organized as follows. Section 2 presents the basic concepts. Section 3 provides the probabilistic fundamentals exploited by the obfuscation operators. Section 4 introduces the basic obfuscation operators used to protect the privacy of the users. Section 5 presents the composition of our basic obfuscation operators and the set of all the available operators. Section 6 analyzes our solution against adversarial attacks aimed at compromising the privacy guaranteed to the users. Section 7 presents an experimental study evaluating the robustness of our solution. Section 8 describes a real application scenario. Section 9 discusses related work. Section 10 presents our conclusions.

2 BASIC CONCEPTS

The physical position of users, as each physical measurement, is always affected by an intrinsic measurement error introduced by sensing technologies. A direct consequence of such a lack of precision is that the location position of a user cannot be expressed as a geographical point, which would imply to suppose that sensing technologies can return exact information.¹ We then assume that positions of users are always represented as *planar circular areas*. This assumption satisfies the general requirement of considering convex areas to easily compute integrals over them. Also, circular areas approximate well the actual shape resulting from many location techniques (e.g., location gathering based on

1. Some works (e.g., [7], [14], [19], [27]) approximate positions as geographic points, which is acceptable when the purpose is to analyze techniques that are affected by small measurement errors only. In general, such an assumption is not realistic since location measurement errors are often a relevant factor of the measurement accuracy.

cellular phones). A *location measurement* returned by a sensing technology can then be defined as follows.

Definition 2.1 (Location measurement): Let (x_u, y_u) be the real position of a user u . A *location measurement* for u is a circular area $A_i = \langle x_i, y_i, r_i \rangle \subseteq \mathbb{R}^2$ returned by a sensing technology such that (x_i, y_i) are the coordinates of the center of A_i , r_i is its radius, and the following conditions hold:

- 1) $P((x_u, y_u) \in A_i) = 1$;
- 2) $P((x_u, y_u) \in A)$, where $A = \langle x, y, \delta r \rangle \subset A_i$ is the neighborhood of position (x, y) with δr an infinitely small radius, is uniformly distributed.

Condition 1 comes from observing that sensing technologies based on cellular phones usually guarantee that the real user position is within the returned area [12]. Condition 2 states that the probability that the real user position falls within a neighborhood $A \subset A_i$ of a random point (x, y) is uniformly distributed. In other words, the real user position could be randomly located everywhere inside A_i with uniform probability.

The goal of our work is to design a solution that protects the location privacy of the users according to their preferences and application context. To this end, the location privacy must be measured and quantified with respect to the *accuracy* of the location measurement: the more accurate the measurement, the less the privacy. The accuracy of a location measurement returned by a sensing technology depends on the radius of the measured circular area, which, in turn, depends on the unavoidable measurement error of the sensing technology. To evaluate the quality of a given location measurement, its accuracy must then be compared with the best accuracy that sensing technologies are able to provide. Several works describe and discuss different location techniques and their best accuracy [20], [32], which is always expressed by defining the radius of the area returned if the best accuracy is achieved.

We introduce a metric, called *relevance*, that provides both an adimensional technology-independent measure of the location accuracy and a measure of the privacy of a location measurement. The relevance associated with a location measurement is formally defined as follows.

Definition 2.2 (Relevance): Let $A_i = \langle x_i, y_i, r_i \rangle$ be a location measurement for a user and r_o be the radius of the area that would be produced if the optimal accuracy is achieved. The *relevance* associated with A_i , denoted \mathcal{R}_i , is the ratio r_o^2/r_i^2 .

In other words, \mathcal{R}_i models the relative accuracy loss of a given measure (e.g., due to particular environmental conditions) with respect to the optimal accuracy r_o that the location techniques would have achieved in perfect environmental conditions. \mathcal{R}_i is the only relevance value that depends on physical values (i.e., measurement errors). By definition, such a relevance:

- tends to 0, when the location measurement is extremely inaccurate;

- is equal to 1, when the location measurement has achieved the best accuracy that the location techniques allow;
- is in the range (0,1), otherwise; the higher the value, the higher the accuracy.

The *location privacy* associated with a location measurement A_i can then be defined as follows.

Definition 2.3 (Location privacy): Let A_i be a location measurement with relevance \mathcal{R}_i . The *location privacy* of A_i is $1 - \mathcal{R}_i$.

In our reference scenario, users can specify their privacy preferences in term of a *final relevance* \mathcal{R}_f that a location measurement must not exceed. A typical way to let users specify their privacy preferences, which has been presented in the literature (e.g., [5], [14]), is based on the concept of *minimum distance*. For instance, a user can define “100 meters” as her privacy preference, meaning that she can be located with an accuracy not better than 100 meters. Considering measurements that produce circular areas, such a preference corresponds to an area of radius 100 meters at least. Although this solution is certainly intuitive and easily understandable by users, it suffers from some drawbacks. In particular, a minimum distance is meaningful in a specific application context only, and is suitable when the obfuscation is performed by scaling a location measurement to a coarser granularity. We instead propose a solution based on the specification of a final relevance \mathcal{R}_f that does not depend on the application context and provides strong robustness. The final relevance \mathcal{R}_f together with the initial relevance \mathcal{R}_i associated with A_i are used to derive the *accuracy degradation* that needs to be introduced for privacy reason.

Definition 2.4 (Accuracy degradation): Let A_i be a location measurement with initial relevance \mathcal{R}_i , and let \mathcal{R}_f be the final relevance requested by the user. The *accuracy degradation* to be applied to A_i , denoted λ , is the ratio $\mathcal{R}_f / \mathcal{R}_i$.

Given a location measurement and an accuracy degradation, our problem is to transform (*obfuscate*) the location measurement in such a way that the resulting area satisfies the privacy preference \mathcal{R}_f defined by the user.

Problem 2.1 (Obfuscation): Let (x_u, y_u) be the real position of a user u , A_i with relevance \mathcal{R}_i be a location measurement for u , and \mathcal{R}_f be the final relevance to be satisfied. Transform A_i into an obfuscated area A_f such that the following conditions hold:

- 1) A_f has relevance \mathcal{R}_f ;
- 2) $P((x_u, y_u) \in A_f) > 0$.

Condition 1 requires the obfuscated area to satisfy the privacy preference of the user. Condition 2 requires the obfuscated area to include the real user position, and implies that A_i and A_f cannot be disjoint.

The transformation of a location measurement A_i into an obfuscated area A_f is performed by applying a set

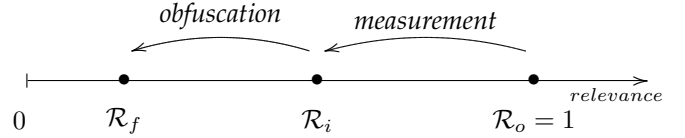


Fig. 1. Relevance degradation due to the intrinsic measurement error and obfuscation

of basic *obfuscation operators* (or a combination of them) that change the radius, or the center, of the original location measurement. As illustrated in Figure 1, the transformation of A_i into A_f introduces a relevance degradation in addition to the natural degradation due to the intrinsic measurement error. Note that if $\mathcal{R}_f \geq \mathcal{R}_i$, no obfuscation is applied to the location measurement, since the measurement error introduced by a sensing technology already satisfies the privacy preference of the user. The following sections describe the basic obfuscation operators and their composition.

3 PROBABILISTIC FUNDAMENTALS OF THE OBFUSCATION OPERATORS

We briefly survey the basic probabilistic concepts exploited by our obfuscation operators.

Considering the two coordinates (x, y) as two random variables, Definition 2.1 implies that each location measurement is characterized by a *joint probability density function* (joint pdf) that is uniform within the location measurement itself [28].

Definition 3.1 (Joint pdf): Given a location measurement $A_i = \langle x_i, y_i, r_i \rangle$, the *joint probability density function* (joint pdf) of variables X, Y corresponding to the x-coordinate and the y-coordinate, respectively, denoted $f_i(X, Y)$, is:

$$f_i(x, y) = \begin{cases} \frac{1}{\pi r_i^2} & \text{if } (x, y) \in A_i \\ 0 & \text{otherwise.} \end{cases}$$

The corresponding joint cumulative distribution function (joint cdf) F_i computed over the location measurement A_i (i.e., $\int \int_{A_i} f_i(x, y) dx dy$) is equal to 1. Intuitively, the joint pdf represents the probability distribution of the real user position to be in the neighborhood of a point $(x, y) \in A_i$; the joint cdf over A_i is the probability that the real user position is within A_i . The physical transformations that can be applied on A_i , that is, a change in its radius or center, produce an obfuscated area A_f for which the joint pdf, joint cdf, or both may be different from the joint pdf and the joint cdf of the original location measurement. Such physical transformations introduce in the original location measurement an accuracy degradation λ (see Definition 2.4) that can be defined as the composite probability of the following two independent events: *i*) a random point $(x', y') \in A_f$ belongs to the intersection between A_i and A_f , and *ii*) the user’s actual position (x_u, y_u) belongs to the intersection. The term λ is then equal to:

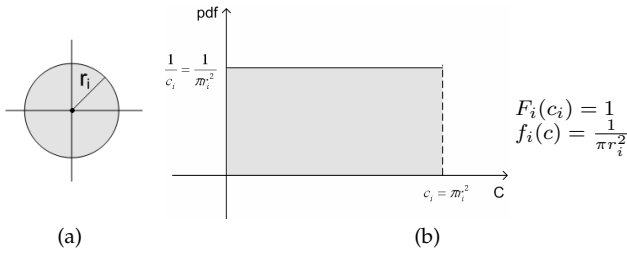


Fig. 2. A location measurement (a) and the pdf of the corresponding variable C (b)

$$\begin{aligned} \lambda &= P((x', y') \in (A_i \cap A_f)) \cdot P((x_u, y_u) \in (A_i \cap A_f)) = \\ &= \frac{(A_i \cap A_f)}{A_f} \frac{(A_i \cap A_f)}{A_i} = \frac{(A_i \cap A_f)^2}{A_f A_i} \end{aligned} \quad (1)$$

From Definition 2.4 and Equation (1) we obtain that:

$$\lambda = \frac{\mathcal{R}_f}{\mathcal{R}_i} = \frac{(A_i \cap A_f)^2}{A_f A_i} \quad (2)$$

Equation 2 represents the relationship between the accuracy degradation λ and the original location measurement A_i , which are known, and the corresponding obfuscated area A_f , which needs to be computed.

In the following, to graphically illustrate the probabilistic effects of an obfuscation over a location measurement A_i , we consider a continuous random variable C , defined on the nonnegative real numbers, with a uniform distribution on $[0, \pi r_i^2]$, meaning that the probability density function $f_i(c) = \frac{1}{\pi r_i^2}$, $c \in [0, \pi r_i^2]$ (see Figure 2). The corresponding cumulative distributed function F_i computed for $c_i = \pi r_i^2$, which is the gray area under the pdf from 0 to πr_i^2 in Figure 2, is equal to 1. It easy to see that the random variable C is statistically equivalent to variables (X, Y) .

4 BASIC OBFUSCATION OPERATORS

An obfuscation operator calculates an obfuscated area A_f with relevance \mathcal{R}_f , starting from a location measurement A_i with relevance \mathcal{R}_i . Formally, an obfuscation operator is defined as follows.

Definition 4.1 (Obfuscation operator): Let \mathcal{A} be the set of circular areas. An obfuscation operator $\text{op}: \mathcal{A} \times (0, 1] \times (0, 1] \rightarrow \mathcal{A}$ takes a circular area A_i and two relevance values \mathcal{R}_i and \mathcal{R}_f as input, where \mathcal{R}_i is the relevance associated with A_i and $\mathcal{R}_f < \mathcal{R}_i$ is the final relevance to be satisfied, and produces as output an obfuscated area A_f such that:

- 1) A_f has relevance \mathcal{R}_f ;
- 2) $A_f \cap A_i \neq \emptyset$.

Here, Condition 2 directly derives from Condition 2 of Problem 2.1, which requires that each obfuscated area has a probability greater than zero of containing the real position of the user.

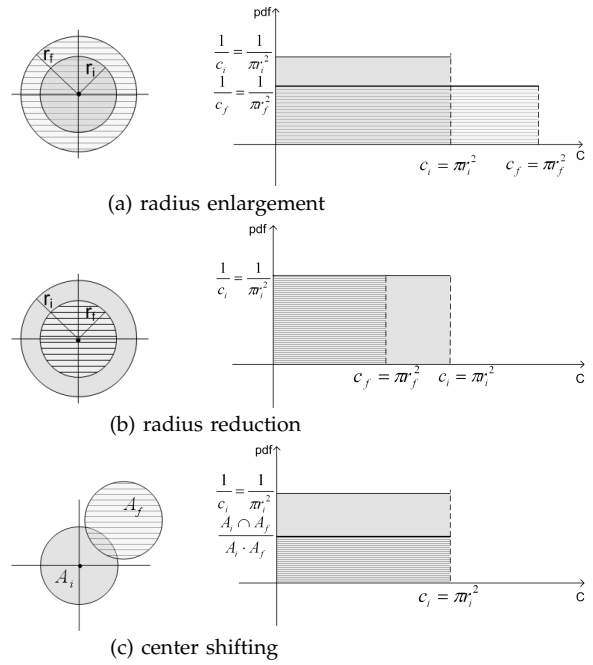


Fig. 3. Graphical illustration of the basic obfuscation operators and of their probabilistic effects on variable C

We now describe our basic obfuscation operators: *enlarge* (E), *reduce* (R), and *shift* (S).

Enlarge (E). Given a location measurement A_i with relevance \mathcal{R}_i , and a relevance \mathcal{R}_f , it produces an obfuscated area $\text{E}(A_i, \mathcal{R}_i, \mathcal{R}_f) = A_f$ with radius $r_f > r_i$ (see Figure 3(a)). Obfuscating a location measurement by increasing its radius logically corresponds to generalization techniques employed in data privacy solutions (e.g., [11]). Such an obfuscation has the effect of decreasing the probability that the real user position falls within the neighborhood of a point $(x, y) \in A_f$, which corresponds to decreasing the pdf's value associated with A_f , while the probability that the real user position falls within A_f remains equal to 1. Considering variable C , Figure 3(a) shows that by enlarging the radius, the pdf's value associated with A_f decreases (from $f_i(c) = \frac{1}{\pi r_i^2}$ to $f_f(c) = \frac{1}{\pi r_f^2}$) while the interval on which is defined increases (from $[0, \pi r_i^2]$ to $[0, \pi r_f^2]$), thus maintaining the area under the pdf equal to 1 (i.e., $F_f(c_f) = F_i(c_i) = 1$).

From Equation (2), it follows that:

$$\frac{\mathcal{R}_f}{\mathcal{R}_i} = \frac{(A_i \cap A_f)^2}{A_f A_i} = \frac{A_i}{A_f} = \frac{r_i^2}{r_f^2} \quad (3)$$

Consequently, the radius r_f of the obfuscated area calculated with this operator satisfying the user privacy preference \mathcal{R}_f is:

$$r_f = r_i \sqrt{\frac{\mathcal{R}_i}{\mathcal{R}_f}}. \quad (4)$$

For instance, suppose that the privacy preference of the user is $\mathcal{R}_f = 0.16$ and that a location measurement with best accuracy has radius $r_o = 0.4$ km (this value

is far from reality, but it is assumed for simplicity). Consider a location measurement A_i with radius $r_i = 0.5$ km. The relevance \mathcal{R}_i associated with A_i is $\mathcal{R}_i = \frac{r_o^2}{r_i^2} = 0.64$. The application of operator \mathbb{E} produces an obfuscated area with relevance \mathcal{R}_f and radius $r_f = r_i \sqrt{\frac{\mathcal{R}_i}{\mathcal{R}_f}} = 1$ km.

Reduce (R). Given a location measurement A_i with relevance \mathcal{R}_i , and a relevance \mathcal{R}_f , it produces an obfuscated area $\mathbb{R}(A_i, \mathcal{R}_i, \mathcal{R}_f) = A_f$ with radius $r_f < r_i$ (see Figure 3(b)). While this obfuscation effect might appear counterintuitive at first sight, it has a precise probabilistic explanation: the probability that the real user position falls within the obfuscated area is reduced, which corresponds to decreasing the area under the pdf associated with A_f , while the pdf's value associated with A_f remains unchanged (i.e., $f_f(c) = f_i(c) = \frac{1}{\pi r_f^2}$). Considering variable C , Figure 3(b) shows that by reducing the radius, the interval on which the pdf associated with A_f is defined decreases (from $[0, \pi r_i^2]$ to $[0, \pi r_f^2]$), meaning that the area under the pdf decreases (i.e., $F_f(c_f) < F_i(c_i) = 1$).²

Equation (2) is again used to compute the radius r_f of the obfuscated area calculated with this technique and that satisfies the user privacy preference \mathcal{R}_f :

$$\frac{\mathcal{R}_f}{\mathcal{R}_i} = \frac{(A_i \cap A_f)^2}{A_f A_i} = \frac{A_f}{A_i} = \frac{r_f^2}{r_i^2} \quad (5)$$

$$r_f = r_i \sqrt{\frac{\mathcal{R}_f}{\mathcal{R}_i}}. \quad (6)$$

For instance, suppose that the privacy preference of the user is $\mathcal{R}_f = 0.16$ and that a location measurement with best accuracy has radius $r_o = 0.4$ km. Consider a location measurement A_i with radius $r_i = 0.5$ km. The relevance \mathcal{R}_i associated with A_i is $\mathcal{R}_i = \frac{r_o^2}{r_i^2} = 0.64$. The application of operator \mathbb{R} produces an obfuscated area with relevance \mathcal{R}_f and radius $r_f = r_i \sqrt{\frac{\mathcal{R}_f}{\mathcal{R}_i}} = 0.25$ km.

Shift (S). Given a location measurement A_i with relevance \mathcal{R}_i , and a relevance \mathcal{R}_f , it produces an obfuscated area $\mathbb{S}(A_i, \mathcal{R}_i, \mathcal{R}_f) = A_f$ such that $(x_f, y_f) = (x_i + d \sin \theta, y_i + d \cos \theta)$, where $d \in (0, 2r_i]$ is the distance between the centers of A_i and of A_f , and $r_f = r_i$ (see Figure 3(c)). Note that distance d cannot be greater than $2r_i$, since by Definition 4.1 the two areas cannot be disjoint. Such an obfuscation has the probabilistic effect of decreasing both the probability that the real user position is in the neighborhood of a point $(x, y) \in A_f$ and the probability that the real user position falls within A_f . Considering variable C , Figure 3(c) shows that by shifting the center, the pdf's value associated with A_f decreases (i.e., $f_f(c) < f_i(c)$) while the interval on which

it is defined remains unchanged, meaning that the area under the pdf decreases (i.e., $F_f(c_f) < F_i(c_i) = 1$). With respect to data privacy literature, it logically corresponds to inserting random noise into the data (e.g., [11]).

With shifting, the obfuscation depends on the intersection of A_i and A_f : the smaller the intersection (i.e., the higher the d), the highest the obfuscation. In particular, the maximum privacy is obtained for $d = 2r_i$. In addition to distance d , a rotation angle θ must be specified to derive an obfuscated area by shifting the center. For the scope of this paper, and without loss of generality, θ is assumed to be randomly generated. Strategies for selecting a value for θ depend on the application context [2]. From Equation (2) and since A_i and A_f have the same area (i.e., $\pi r_i^2 = \pi r_f^2$), it follows that:

$$A_i \cap A_f = \pi r_i^2 \cdot \sqrt{\frac{\mathcal{R}_f}{\mathcal{R}_i}}. \quad (7)$$

Expanding the term $A_i \cap A_f$ as a function of the distance d between the centers, distance d can be calculated numerically by solving the following system of equations, where σ and γ are the central angles of the circular sectors identified by the two radii connecting the center of A_i and of A_f with the intersection points of A_i and of A_f , and $\lambda = \frac{\mathcal{R}_f}{\mathcal{R}_i}$ represents the accuracy degradation.

$$\begin{cases} \left[\frac{\sigma}{2} r_i^2 - \frac{r_i^2}{2} \sin \sigma \right] + \left[\frac{\gamma}{2} r_f^2 - \frac{r_f^2}{2} \sin \gamma \right] = \sqrt{\lambda} \pi r_i r_f \\ d = r_i \cos \frac{\sigma}{2} + r_f \cos \frac{\gamma}{2} \\ r_i \sin \frac{\sigma}{2} = r_f \sin \frac{\gamma}{2} \end{cases} \quad (8)$$

To calculate the distance d between the centers of two partially overlapped circles having the same radius (i.e., $r_i = r_f$), the previous system of equations is simplified as follows.

$$\begin{cases} \sigma - \sin \sigma = \sqrt{\lambda} \pi \\ d = 2r_i \cos \frac{\sigma}{2} \end{cases} \quad (9)$$

Given distance d and a random angle θ , the resulting obfuscated area satisfies the privacy preference \mathcal{R}_f of the user.

For instance, suppose that the privacy preference of the user is $\mathcal{R}_f = 0.4$ and that a location measurement with best accuracy has radius $r_o = 0.895$ km. Consider a location measurement A_i with radius $r_i = 1$ km. The relevance \mathcal{R}_i associated with A_i is $\mathcal{R}_i = \frac{r_o^2}{r_i^2} = 0.8$. By Equation 9, the application of operator \mathbb{S} produces an obfuscated area with relevance \mathcal{R}_f and such that $d = 0.464$ km is the distance between the centers of the two areas. Finally, an angle θ is selected and the obfuscated area is generated.

Figure 4 summarizes the three basic obfuscation operators, along with their input parameters, and shows how an obfuscated area A_f is computed, by reporting the coordinate of its center and the radius.

² Obfuscation by radius reduction, while always suitable in theory, has an obvious limitation in the actual size of location measurements. For instance, GPS locations, being usually affected by small measurement errors, are unsuitable for this technique while cellular phones or wi-fi location measurements may exhibit measurement errors that make reduction applicable, especially if combined with shifting, as discussed in the following.

Operator	Obfuscated area $A_f = \langle x_f, y_f, r_f \rangle$			Comment
	x_f	y_f	r_f	
$\mathbb{E}(A_i, \mathcal{R}_i, \mathcal{R}_f)$	x_i	y_i	$r_i \sqrt{\frac{\mathcal{R}_i}{\mathcal{R}_f}}$	r_f from Eq. (4)
$\mathbb{R}(A_i, \mathcal{R}_i, \mathcal{R}_f)$	x_i	y_i	$r_i \sqrt{\frac{\mathcal{R}_f}{\mathcal{R}_i}}$	r_f from Eq. (6)
$\mathbb{S}(A_i, \mathcal{R}_i, \mathcal{R}_f)$	$x_i + d \sin \theta$	$y_i + d \cos \theta$	r_i	d from Eq. (9) θ random

Fig. 4. Basic obfuscation operators

5 COMPOSITION OF THE BASIC OBFUSCATION OPERATORS

The basic obfuscation operators just illustrated transform a location measurement by changing its radius (operators \mathbb{E} and \mathbb{R}) or by changing its center (operator \mathbb{S}). These two types of physical transformations can also be applied together, meaning that the basic operators can be composed by executing them in sequence. In this case, each operator used in the composition must produce an area where the relevance degradation is always evaluated with respect to the original location measurement A_i and relevance \mathcal{R}_i , which we call *reference area* and *reference relevance*, respectively. This observation changes the definition of obfuscation operator as follows.

Definition 5.1 (Obfuscation operator): Let \mathcal{A} be the set of circular areas, and $A_i \in \mathcal{A}$ be the reference area with reference relevance \mathcal{R}_i . An obfuscation operator $\text{op}_{A_i, \mathcal{R}_i} : \mathcal{A} \times (0, 1] \rightarrow \mathcal{A}$ over A_i and \mathcal{R}_i takes an area A and a relevance \mathcal{R}' as input, with $A \cap A_i \neq \emptyset$ and $\mathcal{R}' < \mathcal{R}_i$, and produces an obfuscated area A' as output such that:

- 1) A' has relevance \mathcal{R}' ;
- 2) $A' \cap A_i \neq \emptyset$.

From Definition 5.1, it follows that two obfuscation operators can be composed only if they are defined and evaluated over the same reference area A_i with reference relevance \mathcal{R}_i . From Definition 4.1 and Definition 5.1, it also follows that $\text{op}_{A_i, \mathcal{R}_i}(A_i, \mathcal{R}_f) \equiv \text{op}_{A_i, \mathcal{R}_i}(A_i, \mathcal{R}_f)$, meaning that when the reference area A_i is also the area that needs to be obfuscated, the two operator definitions are equivalent. In the following, the composition of two obfuscation operators h_{A_i, \mathcal{R}_i} and k_{A_i, \mathcal{R}_i} , called *composed obfuscation operator*, is denoted hk (omitting both the reference area A_i and the reference relevance \mathcal{R}_i) and states that the application of operator h is followed by the application of operator k . As an example, consider composed operator $\text{ES} = \mathbb{S}_{A_i, \mathcal{R}_i}(\mathbb{E}_{A_i, \mathcal{R}_i}(A_i, \mathcal{R}_m), \mathcal{R}_f)$ illustrated in Figure 5, where A_i is the original location measurement (the dark gray area), A_m is the obfuscated area produced by the first operator (the area filled with vertical lines), A_f is the obfuscated area produced by the second operator (the area filled with horizontal lines). In the first obfuscation step (\mathbb{E}), relevance \mathcal{R}_m is a random value between \mathcal{R}_f and \mathcal{R}_i and radius r_m of area A_m is computed as $r_m = r_i \sqrt{\frac{\mathcal{R}_i}{\mathcal{R}_m}}$ (Equation (4)). According

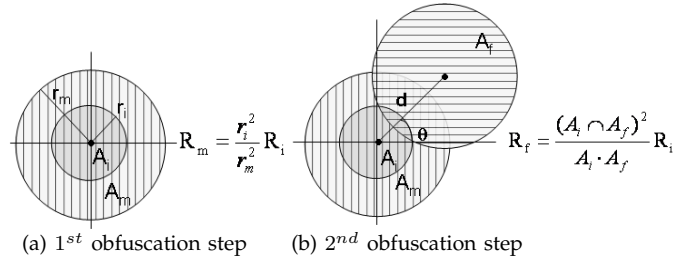
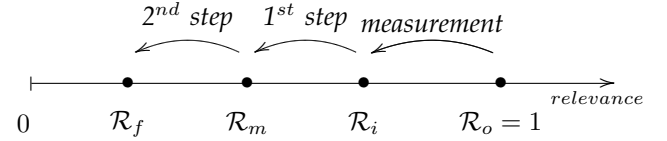
Fig. 5. Composed operator ES applied on A_i 

Fig. 6. Relevance degradation due to intrinsic measurement error and obfuscation

to Definition 5.1, in the second obfuscation step the codomain of the Shift operator must be restricted to the areas that have an intersection with A_i . The value d of the center-shifting is then calculated, starting from area A_m , to generate an obfuscated area A_f whose overlap with the original area A_i satisfies the privacy preference of the user.

Although in theory it is possible to combine operators \mathbb{E} , \mathbb{R} , and \mathbb{S} an arbitrary number of times, the combination of more than two operators is never necessary, as stated by the following lemma.

Lemma 5.1: Given $A_1 = \langle x_1, y_1, r_1 \rangle$ and $A_2 = \langle x_2, y_2, r_2 \rangle$, A_1 can always be transformed into A_2 (or vice versa) by applying one or both (in some order) of these two operations:

- a *center shifting* such that the center (x_1, y_1) of A_1 becomes equal to (x_2, y_2) ;
- a *radius enlargement or reduction* such that r_1 becomes equal to r_2 .

The proof immediately follows from the geometric properties of the circular areas.

From this lemma, it follows that the relevant composed operators are those obtained by combining operators \mathbb{E} and \mathbb{R} with operator \mathbb{S} , that is: ES , SE , RS , and SR . This implies that we only need one intermediate relevance \mathcal{R}_m such that $\mathcal{R}_f < \mathcal{R}_m < \mathcal{R}_i$, which represents the relevance achieved by the first obfuscation step (see Figure 6).

Note that the difference between \mathcal{R}_i and \mathcal{R}_m and the difference between \mathcal{R}_m and \mathcal{R}_f have an impact on the importance associated with each basic operator used in the composition. Indeed, if the difference between the relevances associated with two areas is small, also the corresponding obfuscation effect is small. Figure 7 illustrates the redefinition of the three basic obfuscation operators according to Definition 5.1 and shows the resulting obfuscated area A when they are used: *i*) in the first step (1^{st}) of a composed operator to produce,

	Operator	Obfuscated area $A=\langle x,y,r \rangle$			Comment
		x	y	r	
1 st	$E_{A_i, \mathcal{R}_i}(A_i, \mathcal{R}_m)$	x_i	y_i	$r_i \sqrt{\frac{\mathcal{R}_i}{\mathcal{R}_m}}$	r from Eq. (4)
	$R_{A_i, \mathcal{R}_i}(A_i, \mathcal{R}_m)$	x_i	y_i	$r_i \sqrt{\frac{\mathcal{R}_m}{\mathcal{R}_i}}$	r from Eq. (6)
	$S_{A_i, \mathcal{R}_i}(A_i, \mathcal{R}_m)$	$x_i + d \sin \theta$	$y_i + d \cos \theta$	r_i	d from Eq. (9) θ random
2 nd	$E_{A_i, \mathcal{R}_i}(A_m, \mathcal{R}_f)$	x_m	y_m	$> r_m$	r from Eq. (8)
	$R_{A_i, \mathcal{R}_i}(A_m, \mathcal{R}_f)$	x_m	y_m	$< r_m$	r from Eq. (8)
	$S_{A_i, \mathcal{R}_i}(A_m, \mathcal{R}_f)$	$x_m + d \sin \theta$	$y_m + d \cos \theta$	r_m	d from Eq. (8) θ random

Fig. 7. Redefinition of the obfuscation operators

starting from the original location measurement A_i , an intermediate area with relevance \mathcal{R}_m ; ii) in the second step (2nd) of a composed operator to produce, starting from an intermediate area A_m , the final obfuscated area with relevance \mathcal{R}_f .

Let $\mathcal{A}_{A_i, \mathcal{R}_i}$ be the set of all possible obfuscated areas generated by the application over area A_i with relevance \mathcal{R}_i of the basic and of all composed operators (i.e., ES, SE, RS, and SR). We are interested in finding the set \mathcal{O} of (basic and composed) obfuscation operators that is *complete* and *minimal*, as introduced by the following definition.

Definition 5.2 (Complete and minimal): Given a set \mathcal{O} of obfuscation operators and the set $\mathcal{A}_{A_i, \mathcal{R}_i}^{\mathcal{O}}$ of areas generated by applying any obfuscation operator in \mathcal{O} over a reference area A_i with relevance \mathcal{R}_i , \mathcal{O} is said to be *complete* and *minimal* iff:

- $\mathcal{A}_{A_i, \mathcal{R}_i}^{\mathcal{O}} = \mathcal{A}_{A_i, \mathcal{R}_i}$ (completeness);
- $\forall \mathcal{O}' \subset \mathcal{O}, \exists A' \in \mathcal{A}_{A_i, \mathcal{R}_i}^{\mathcal{O}'} : A' \notin \mathcal{A}_{A_i, \mathcal{R}_i}^{\mathcal{O}}$ (minimality).

A set \mathcal{O} of obfuscation operators is then complete and minimal when it can produce every possible obfuscated area and therefore does not exist another set \mathcal{O}' of obfuscation operators that can produce every possible obfuscated area and that is a proper subset of \mathcal{O} . To determine a complete and minimal set of obfuscation operators, it is important to note that the order in which operators are applied affects the set of areas that can be produced, as showed by the following lemma.

Lemma 5.2: Let A_i with relevance \mathcal{R}_i be the reference area. Given composed operators SE, ES, SR, and RS, the sets of areas that can be produced by applying them over A_i satisfy the following relationships: 1) $\mathcal{A}_{A_i, \mathcal{R}_i}^{SE} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{ES}$; 2) $\mathcal{A}_{A_i, \mathcal{R}_i}^{SR} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{RS}$; 3) $\mathcal{A}_{A_i, \mathcal{R}_i}^{ES} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{SE}$; and 4) $\mathcal{A}_{A_i, \mathcal{R}_i}^{RS} \subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{SR}$.

Proof:

- 1) $\mathcal{A}_{A_i, \mathcal{R}_i}^{SE} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{ES}$. Let $A_f \in \mathcal{A}_{A_i, \mathcal{R}_i}^{SE}$ be an obfuscated area such that A_f contains the original area A_i . A_f can never be produced by operator ES. As a matter of fact, in operator ES, the first step (enlargement) would produce an area A_m that necessarily includes A_i . From Lemma 5.1, A_m has the same radius as A_f and therefore, by definition (Equation 2) has the same relevance as A_f . Since

each step of a composed operator must decrease the relevance (Definition 5.1), A_f can never be returned by the second (shifting) step.

- 2) $\mathcal{A}_{A_i, \mathcal{R}_i}^{SR} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{RS}$. Let $A_f \in \mathcal{A}_{A_i, \mathcal{R}_i}^{SR}$ be an obfuscated area included in the original area A_i . The proof is analogous to case 1 above as reduction applied as a first step would produce an area A_m included in A_i and therefore with same relevance as A_f , which therefore could never be returned by the second (shifting) step.
- 3) $\mathcal{A}_{A_i, \mathcal{R}_i}^{ES} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{SE}$. Let $A_f \in \mathcal{A}_{A_i, \mathcal{R}_i}^{ES}$ be an obfuscated area such that the distance d between the center of A_i and the center of A_f is greater than $2r_i$. A_f can never be produced by operator SE. As a matter of fact, to produce A_f with operator SE, the first step (shifting) would have to produce an area A_m that has empty intersection with original area A_i ; this is not possible by definition (Definition 5.1, Condition 2).
- 4) $\mathcal{A}_{A_i, \mathcal{R}_i}^{RS} \subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{SR}$. It is easy to see that $\forall A_f \in \mathcal{A}_{A_i, \mathcal{R}_i}^{RS}$ with relevance $\mathcal{R}_f < \mathcal{R}_i$, A_f is always partially overlapped with A_i , and the distance d between the center of A_i and the center of A_f is less than or equal to $r_i + r_f$. This implies that $\forall A_f \in \mathcal{A}_{A_i, \mathcal{R}_i}^{RS}$, A_f can also be obtained by first shifting A_i , thus obtaining an area A_m with $(x_m, y_m) = (x_f, y_f)$ and $\mathcal{R}_m < \mathcal{R}_i$, and then by reducing the radius of A_m until r_m becomes equal to r_f , thus obtaining an area A_f with relevance $\mathcal{R}_f < \mathcal{R}_m < \mathcal{R}_i$. \square

From Lemma 5.2, we can immediately conclude that composed operator RS is redundant since it can only produce areas that can be produced by composed operator SR. The set $\mathcal{O} = \{E, R, S, ES, SE, SR\}$ of obfuscation operators over A_i and \mathcal{R}_i is then complete and minimal, as captured by the following theorem.

Theorem 5.1: Given a reference area A_i with relevance \mathcal{R}_i , the set $\mathcal{O} = \{E, R, S, ES, SE, SR\}$ of obfuscation operators over A_i and \mathcal{R}_i is complete and minimal.

Proof: The proof follows from Lemmas 5.1 and 5.2. \square

Figure 8 summarizes our composed operators reporting, for each of them, the coordinate of the center and the radius of the intermediate area A_m , computed through the first operator of the composed operator, and the coordinate of the center and the radius of the final obfuscated area A_f , computed through the second operator of the composed operator and satisfying user privacy preference \mathcal{R}_f . Note that for composed operators SE and SR the figure distinguishes two different cases, depending on whether the resulting obfuscated area A_f : 1) is partially overlapped with A_i , 2) is fully included in A_i (for SR), or it fully includes A_i (for SE). The reason for this is that the partial overlapping and inclusion cases must be treated separately, since they have different behaviors when analyzed with respect to the adversary that tries to reduce the obfuscation effects.

Op	Schema	1 st step ($A_m = \langle x_m, y_m, r_m \rangle$)			2 nd step ($A_f = \langle x_f, y_f, r_f \rangle$)			Comment
		x_m	y_m	r_m	x_f	y_f	r_f	
ES		x_i	y_i	$r_i \sqrt{\frac{\mathcal{R}_i}{\mathcal{R}_f}}$	$x_m + d \sin \theta$	$y_m + d \cos \theta$	r_m	d from Eq. (8) θ random
SE	SE partial overlapping 	$x_i + d \sin \theta$	$y_i + d \cos \theta$	r_i	x_m	y_m	$> r_m$	d from Eq. (9) r_f from Eq. (8)
	SE inclusion (particular case) 	$x_i + d \sin \theta$	$y_i + d \cos \theta$	r_i	x_m	y_m	$r_m \sqrt{\frac{\mathcal{R}_i}{\mathcal{R}_f}}$	d from Eq. (9) θ random
SR	SR partial overlapping 	$x_i + d \sin \theta$	$y_i + d \cos \theta$	r_i	x_m	y_m	$< r_m$	d from Eq. (9) r_f from Eq. (8)
	SR inclusion (particular case) 	$x_i + d \sin \theta$	$y_i + d \cos \theta$	r_i	x_m	y_m	$r_m \sqrt{\frac{\mathcal{R}_f}{\mathcal{R}_i}}$	d from Eq. (9) θ random

Fig. 8. Composed operators

6 ADVERSARY MODEL

A sound definition of relevance as a metric for estimating the location accuracy and the privacy is not enough to measure the real privacy protection provided by the obfuscation operators, because the degree of robustness of each operator must be evaluated with respect to possible de-obfuscation attempts adversaries can perform. Accordingly, we say that *an obfuscation operator is robust if and only if it cannot be reversed by an adversary to obtain a location measurement that approximates the original location measurement better than the obfuscated area*, meaning that the relevance associated with the de-obfuscated area is greater than the relevance associated with the obfuscated area. It follows that two issues must be considered when the obfuscation robustness is analyzed:

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

While it is relatively straightforward to de-obfuscate an area by applying some transformations, understanding

whether the de-obfuscated area is more or less accurate than the obfuscated area could be an irresolvable task for the adversary. In this situation, called *blind de-obfuscation*, the adversary can only act randomly and the obfuscation operators that permit just this possibility are considered *strongly robust*. However, we will see that some operators, called *weakly robust*, may provide the adversary with a preferred de-obfuscation strategy.

In our analysis, we assume an *adversary model* where all parties that receive or manage an obfuscated area without knowing the original location measurement are considered untrusted and could behave as adversaries. In addition, we assume that the adversary is aware of *i) the obfuscated area, ii) the location sensing technology adopted by the location service, and iii) all the available obfuscation operators*. The specific obfuscation operators applied to produce the obfuscated area, as well as the relevance of the obfuscated area, are instead assumed to be unknown. Note that, we do not explicitly consider the problem of an adversary that infers location information from subsequent queries of a user location. Intuitively,

our solution offers a degree of protection because, by design, each location measurement is obfuscated by applying a technique randomly chosen among a set of possibly obfuscation techniques. Therefore, the uncertainty is increased for an adversary aiming at inferring information. There is also no obvious way for the adversary to calculate a location that proves better, in term of relevance, with respect to the obfuscated areas. However, an extensive analysis of this case is expected in future works.

We can then consider two different scenarios. In the first scenario, the adversary cannot infer any information from the obfuscated area and therefore she only knows that the area has been produced by using an obfuscation operator belonging to the whole set of available operators, which we call \ast -family= $\{E,R,S,ES,SE,SR\}$. In the second scenario, an adversary can collect some reliable application context information that is exploited to infer whether the obfuscated area has a radius apparently “unusually small”, meaning that the obfuscated area has been computed through set $\{R, SR\}$ of operators, or “unusually large”, meaning that the obfuscated area has been computed through set $\{E, SE, ES\}$ of operators. Given their importance in the analysis, we call these two subsets R-family and E-family, respectively. Note that the adversary cannot recognize whether operator S has been used to produce obfuscated areas. Moreover, operator S introduces a random parameter (the rotation angle θ) that the adversary cannot evaluate. The consequence is that if the adversary tries to de-obfuscate the given obfuscated area through a shifting of the center, it cannot evaluate whether the de-obfuscated area has a relevance greater than the relevance associated with the obfuscated area. Therefore, we assume that the adversary tries to de-obfuscate the observed area by enlarging or reducing its radius only.

The ability to recognize the R-family and the E-family allows an adversary to decide if the de-obfuscation attempts should be based on enlarging or reducing the radius of the obfuscated area, respectively. However, the task of recognizing if an obfuscated area has been produced by an operator of these two families could be costly and time-consuming due to the very nature of location measurements, whose accuracy strongly depends on environmental factors, such as weather conditions or building materials. Such a task, except for evidently abnormal values for the radius, is based on the average measurement errors produced by the specific location technique in the specific area of interest (and possibly in the same measurement conditions). Performing this evaluation implies, in general, the availability of a reliable statistic of measurement errors in the observed area, which can be collected as a result of field tests during different days with different environmental conditions.

In the following analysis, we consider the worst scenario, where an adversary is able to distinguish the family of operators used to produce the obfuscated area.

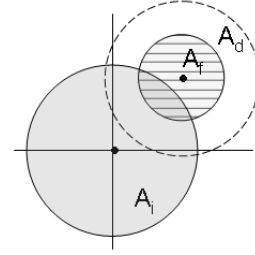


Fig. 9. De-obfuscation attempt on area A_f produced through composed operator SR (partial overlapping)

6.1 R-family de-obfuscation

The R-family de-obfuscation attempts are focused on reversing the obfuscation through an enlargement of the radius of the obfuscated area. As an example, consider the areas reported in Figure 9, where A_i is the original location measurement (the gray area) unknown to the adversary, A_f is the obfuscated area (the area filled with horizontal lines) obtained through operator SR , and A_d is the de-obfuscated area produced by enlarging the radius of A_f (the area with dashed line). We have that $\frac{(A_i \cap A_d) \cdot (A_i \cap A_d)}{A_i \cdot A_d} > \frac{(A_i \cap A_f) \cdot (A_i \cap A_f)}{A_i \cdot A_f}$, meaning that A_d has relevance greater than the relevance associated with A_f (see Equation (7)). For each operator of the R-family, Figure 10 shows the variation of the relevance (Y axis) as a function of the radius of the de-obfuscated area (X axis), where the result of a de-obfuscation attempt is intuitively represented by the $+$ and $-$ labels: a de-obfuscation attempt succeeds when the adversary recovers an area with a relevance greater than the relevance associated with the obfuscated area (label $+$); it fails, otherwise (label $-$). In the analysis, important radii are:

- radius r_f of the obfuscated area, which represents the starting point for a de-obfuscation attempt;
- radius r_{max} of the area with best relevance \mathcal{R}_{max} , which represents the best de-obfuscation that the adversary can achieve;
- radius $r_{i,d}$ of the de-obfuscated area including the original location measurement and intersecting it in a single point. Radius $r_{i,d} = r_i + d$, where r_i is the radius of the original location measurement A_i and d is the distance between the centers of A_i and A_f ;
- radius r_{bp} of an area with the same relevance \mathcal{R}_f associated with the obfuscated area A_f . It represents the breakpoint radius after which the adversary produces a de-obfuscated area of less relevance than \mathcal{R}_f .

From Figure 10, it is easy to see that starting from an obfuscated area A_f with radius r_f , the adversary may increase the relevance (thus decreasing the privacy of the users' locations) by enlarging the radius of the obfuscated area from r_f to r_{bp} . The maximum relevance is obtained for radius r_{max} , then the relevance decreases, while remaining greater than the relevance associated with the obfuscated area until radius r_{bp} is reached.

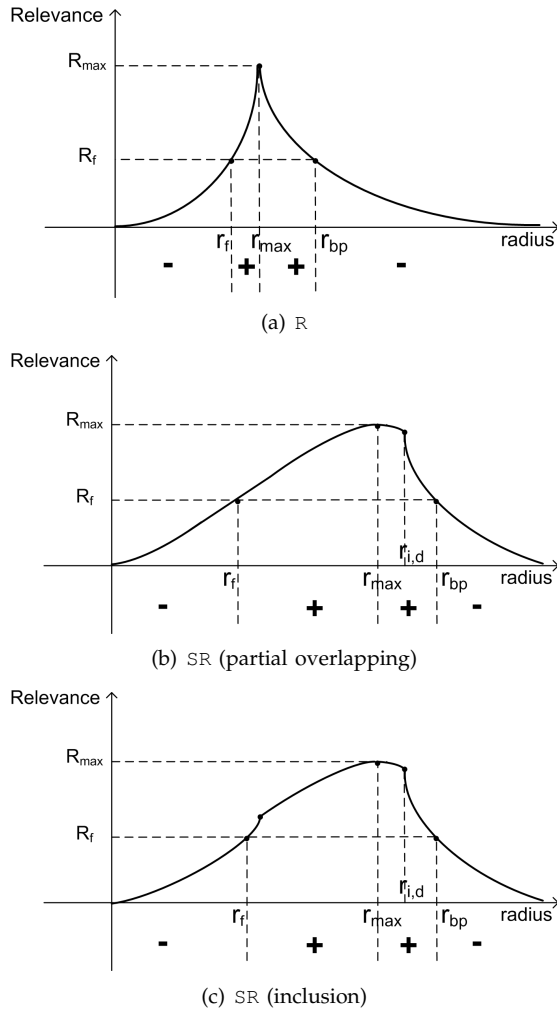


Fig. 10. Relevance variations in de-obfuscation attempts against the R-family

Note that the adversary does not know the values of radii r_{max} and r_{bp} . Furthermore, the curve representing the variation of the relevance (i.e., the *adversary gain*) depends on the specific obfuscation operator used for producing A_f , which is again an information that the adversary does not know. There are therefore three cases. First, if the obfuscated area was produced by operator R, the equations that model the relevance variation (see Figure 10(a)) are based on the quadratic function of operator R (Equation (5)), between r_f and r_{max} , and on the quadratic function of operator E (Equation (3)), between r_{max} and r_{bp} , because the relevance decreases as in the case of an obfuscation produced by enlarging the radius. Radius r_{max} coincides with radius r_i of the original location measurement, which is associated with maximum relevance $\mathcal{R}_{max} = \mathcal{R}_i$.

Second, if the obfuscated area was produced by operator SR (partial overlapping), the equations that model the relevance variation (see Figure 10(b)) are based on the partial overlapping produced by operator S (Equation (7)), between r_f and $r_{i,d}$, and the quadratic function of operator E (Equation (3)), between $r_{i,d}$ and r_{bp} . In this

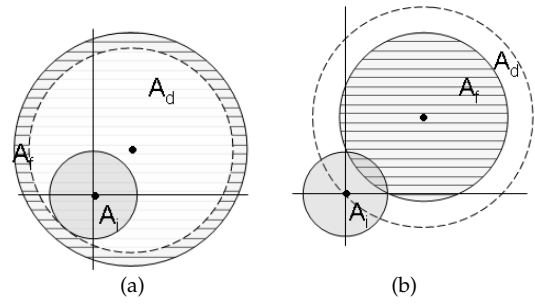


Fig. 11. De-obfuscation attempt on area A_f produced via operator SE (inclusion) (a) and operator ES (b)

case, the maximum relevance \mathcal{R}_{max} that an adversary can achieve in correspondence with radius r_{max} is less than \mathcal{R}_i , since to obtain relevance \mathcal{R}_i , operator S used in the SR process should be de-obfuscated too.

Third, if the obfuscated area was produced by operator SR (inclusion), the only difference with the previous case is that the initial slope of the curve representing the relevance variation (see Figure 10(c)) follows the quadratic function of operator R (Equation (5)).

From this analysis, it follows that a radius enlargement is the strategy that the adversary must apply when an obfuscated area has been produced by an operator of the R-family. For this reason, the R-family exhibits a weak robustness, because if the adversary is able to guess the obfuscation family, she can apply a preferred de-obfuscation strategy based on the enlargement of radius r_f . However, since the adversary is not able to calculate or infer boundary r_{bp} , she can exceed r_{bp} , thus retrieving an area with relevance less than \mathcal{R}_f .

6.2 E-family de-obfuscation

Although it may seem that to de-obfuscate an area produced by an operator of the E-family the adversary should just reduce the radius of the obfuscated area, actually this is not always the case. In fact, to obtain a de-obfuscated area with relevance greater than relevance \mathcal{R}_f , the adversary should try to increase the overlapping between A_i and A_f . If A_i is included in A_f , the adversary should reduce the radius of obfuscated area A_f , while if A_i and A_f are partially overlapped, the adversary should enlarge the radius. To better understand the rationale behind this observation, consider the examples reported in Figure 11, where A_i is the original location measurement, A_f is the obfuscated area, and A_d is the de-obfuscated area produced by manipulating the radius of area A_f . Figure 11(a) illustrates an area A_f obtained through operator SE and such that A_i is included in A_f . In this case, the adversary can actually recover an area having a relevance better than the relevance associated with A_f by reducing the radius of the observed obfuscated area. Considering Equation (3), it follows that $r_i^2/r_d^2 > r_i^2/r_f^2$, meaning that the relevance associated with area A_d is greater than the relevance \mathcal{R}_f associated with area A_f , and then the location privacy of the user decreases.

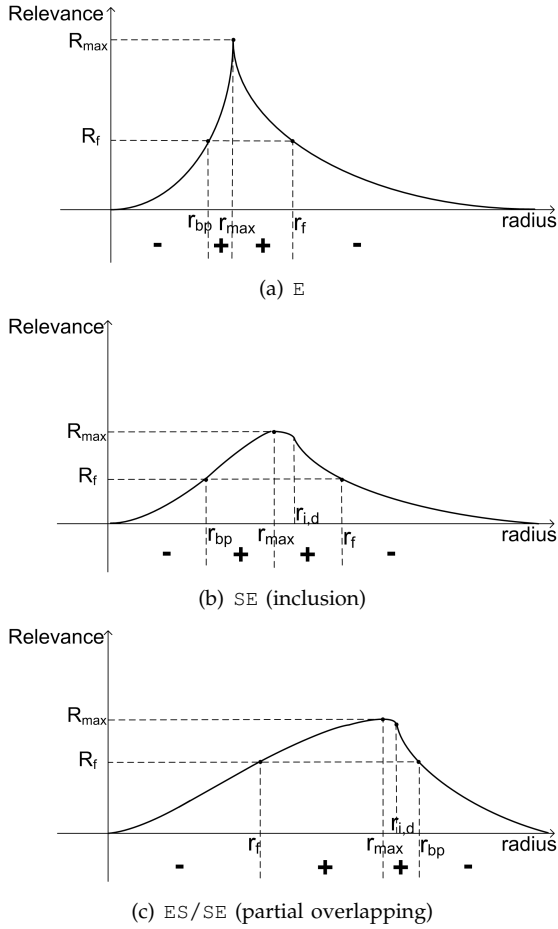


Fig. 12. Relevance variations in de-obfuscation attempts against the E-family

Figure 11(b) illustrates instead an obfuscated area A_f obtained through operator ES and a de-obfuscated area A_d , with a relevance better than the relevance associated with A_f , obtained by enlarging, rather than reducing, the radius of the obfuscated area. Considering Equation (7), we have that $\frac{(A_i \cap A_d) \cdot (A_i \cap A_d)}{A_i \cdot A_d} > \frac{(A_i \cap A_f) \cdot (A_i \cap A_f)}{A_i \cdot A_f}$, meaning that the relevance associated with area A_d is greater than the relevance \mathcal{R}_f associated with area A_f , and again the location privacy of the user decreases.

Like for the R-family, Figure 12 illustrates, for all operators of the E-family, how the relevance varies with respect to a manipulation of the radius of the obfuscated area. Here, again we use radii r_f , r_{max} , $r_{i,d}$, and r_{bp} to denote the radius of the obfuscated area, the radius of the de-obfuscated area with maximal relevance, the radius of the de-obfuscated area including A_i and intersecting A_i in a single point, and the breakpoint radius, respectively. We need to discuss two cases separately.

Case 1: operators E and SE (inclusion) (Figure 12(a) and Figure 12(b)). When the obfuscated area is obtained through operators E and SE (inclusion), the adversary may increase the relevance (thus decreasing the privacy of the users' locations) by *reducing* the radius of the obfuscated area from r_f to r_{bp} . The maximum relevance

is obtained for radius r_{max} , from which the relevance decreases and falls below the relevance associated with the obfuscated area when radius r_{bp} is exceeded. Figure 12(a) shows that for operator E the relevance variation obtained by reducing the radius r_f is modeled by the quadratic function of operator E (Equation (3)), between r_{max} and r_f , and is modeled as an obfuscation produced by reducing the radius (Equation (5)), between r_{bp} and r_{max} . Radius r_{max} coincides with radius r_i of the original location measurement, which is associated with maximum relevance $\mathcal{R}_{max} = \mathcal{R}_i$. Figure 12(b) shows that for operator SE the relevance variation obtained by reducing the radius r_f is again modeled by the quadratic function of operator E (Equation (3)), between $r_{i,d}$ and r_f , and is then modeled as a function of the overlapping of the areas (Equation (7)), between r_{bp} and $r_{i,d}$. A maximum relevance $R_{max} < \mathcal{R}_i$ can be achieved by the adversary with radius r_{max} .

Case 2: operators ES and SE (partial overlapping) (Figure 12(c)). When the obfuscated area is obtained through operators ES and SE (partial overlapping), the adversary may increase the relevance (thus decreasing the privacy of the users' locations) by enlarging the radius of the obfuscated area from r_f to r_{bp} . The maximum relevance is still obtained for radius r_{max} and radius r_{bp} is the breakpoint. Figure 12(c) shows that for these operators the relevance variation obtained by enlarging the radius r_f is modeled by the partial overlapping produced by operator S (Equation (7)), between r_f and $r_{i,d}$, and by the quadratic function of operator E (Equation (3)), between $r_{i,d}$ and r_{bp} .

From our analysis, we can conclude that for the E-family there is not a preferred de-obfuscation strategy. This implies that the adversary is forced to act blindly by randomly choosing a reduction or an enlargement with no information about the outcome. Being the adversary unable to assess the actual relevance gain or loss of the de-obfuscated area with respect to the obfuscated one, the E-family is said to be strongly robust.

6.3 *-family de-obfuscation

The adversary that cannot distinguish between the R-family or the E-family is forced to consider the whole set of available obfuscation operators. According to the previous discussions, an obfuscated area produced through obfuscation operators S, ES, SE (partial overlapping), R, and SR should be de-obfuscated by enlarging its radius, while an obfuscated area produced through obfuscation operators SE (inclusion) and E should be de-obfuscated by reducing its radius. The radius enlargement is then the most likely de-obfuscation strategy for the *-family, although a degree of uncertainty is due to those two operators for which radius reduction would have been the right choice. For this reason, the *-family, in general, shows an intermediate robustness level between the strong one of the E-family and the weak one of the R-family.

7 EXPERIMENTAL STUDY

We experimentally evaluated our obfuscation operators on a dataset of obfuscated areas and by simulating the adversary behavior under different assumptions. During the tests we have measured the robustness of our operators, compared one with the others, and with the trivial solution based on just an enlargement of the location.

7.1 Experimental setup

To build up the datasets of obfuscated locations we produced 20,000 random location measurements, and 20,000 random relevances \mathcal{R}_f to simulate users privacy preferences. We associated each location measurement with a relevance and applied our different obfuscation operators. We produced three different datasets of 20,000 obfuscated areas each produced by applying: 1) the operators belonging to the \mathcal{R} -family, randomly; 2) the operators belonging to the \mathcal{E} -family, randomly; and 3) operator \mathcal{E} only to test the behavior of traditional solutions.

We developed a simulator of the adversary behavior, using *MATLAB 2007a*, which let us apply different de-obfuscation strategies. We considered two main adversary behaviors: *i) no contextual awareness*, when the adversary is not aware of any contextual information and is not able to infer the obfuscation family applied; *ii) contextual awareness*, when the adversary knows enough contextual information to infer the obfuscation family applied. The different assumptions regarding the contextual awareness have consequences on the adversary behaviour that has to be assumed during the evaluation of the \mathcal{R} -family: *i)* the adversary with no contextual information does not know that an operator of the \mathcal{R} -family has been applied, thus she cannot infer that enlarging the obfuscated area is the best strategy. In this case, she will act randomly, either by reducing or enlarging the obfuscated area; *ii)* the adversary that knows that one operator of the \mathcal{R} -family has been applied, will only enlarge the obfuscated area.

For the other two types of obfuscation, the whole \mathcal{E} -family and operator \mathcal{E} , the different adversary behaviors resulting from the contextual awareness are less meaningful. For the \mathcal{E} -family, as we illustrated previously, there is not a preferred strategy, and the adversary will always de-obfuscate randomly by either reducing or enlarging the obfuscated areas, regardless to her contextual awareness. For operator \mathcal{E} the adversary knows that she has an advantage in reducing an obfuscated area.

Finally, we assume that de-obfuscation attempts consist in enlarging/reducing the radius of the obfuscated area by a de-obfuscation level of 10%, 30%, 50%, and 70%. This is aimed to test different adversary behaviors, from the most conservative to the greediest. The hypothesis is that high de-obfuscation levels are associated with both high gains in relevance and low success rates, while low de-obfuscation levels result in low gains and high

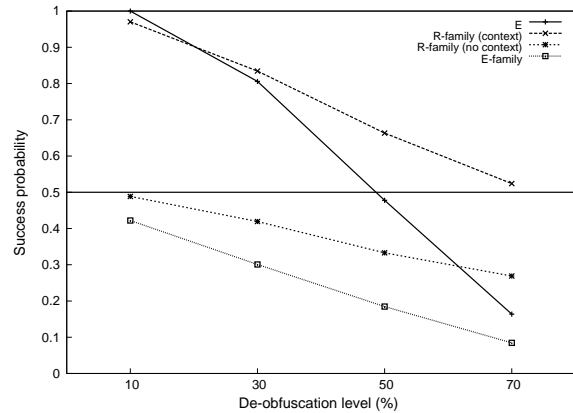


Fig. 13. Rate of successful de-obfuscations attempts based on different degrees of manipulations

success rates. However, we will see that not all operators confirm this hypothesis.

7.2 Experimental results

Quantitative evaluations and comparisons are produced based on both the successful de-obfuscation rate achieved by the adversary and the relevance gain or loss the adversary obtains as a result of a de-obfuscation attempt. Analyzing both aspects (i.e., the success rate and the amount of gain/loss) is relevant because in a real scenario the adversary is assumed to behave strategically by, implicitly or explicitly, maximizing them.

Success rate analysis. A success happens when the resulting de-obfuscated area has a relevance greater than the one associated with the obfuscated area. The dataset produced by applying the operators of the \mathcal{R} -family has been tested twice for the two different adversary behaviors depending on the presence or absence of contextual awareness.

Figure 13 shows how de-obfuscation success rate varies (y-axis) with different levels of de-obfuscation (x-axis), based on the type of obfuscation and contextual awareness. Comparing the four cases, we observe that:

- de-obfuscation attacks against the \mathcal{E} -family and the \mathcal{R} -family with no contextual awareness never exceed a success rate of 50%, which confirms our theoretical result that at best (i.e., for very small radius manipulation) the adversary achieves the same probability of success or failure and she has neither a preferred attack strategy nor the possibility to guess whether the de-obfuscation succeeds;
- de-obfuscation attacks against operator \mathcal{E} and the \mathcal{R} -family with contextual awareness have a success rate ranging from more than 95% for very small radius modification to 80% for a de-obfuscation of 30%;

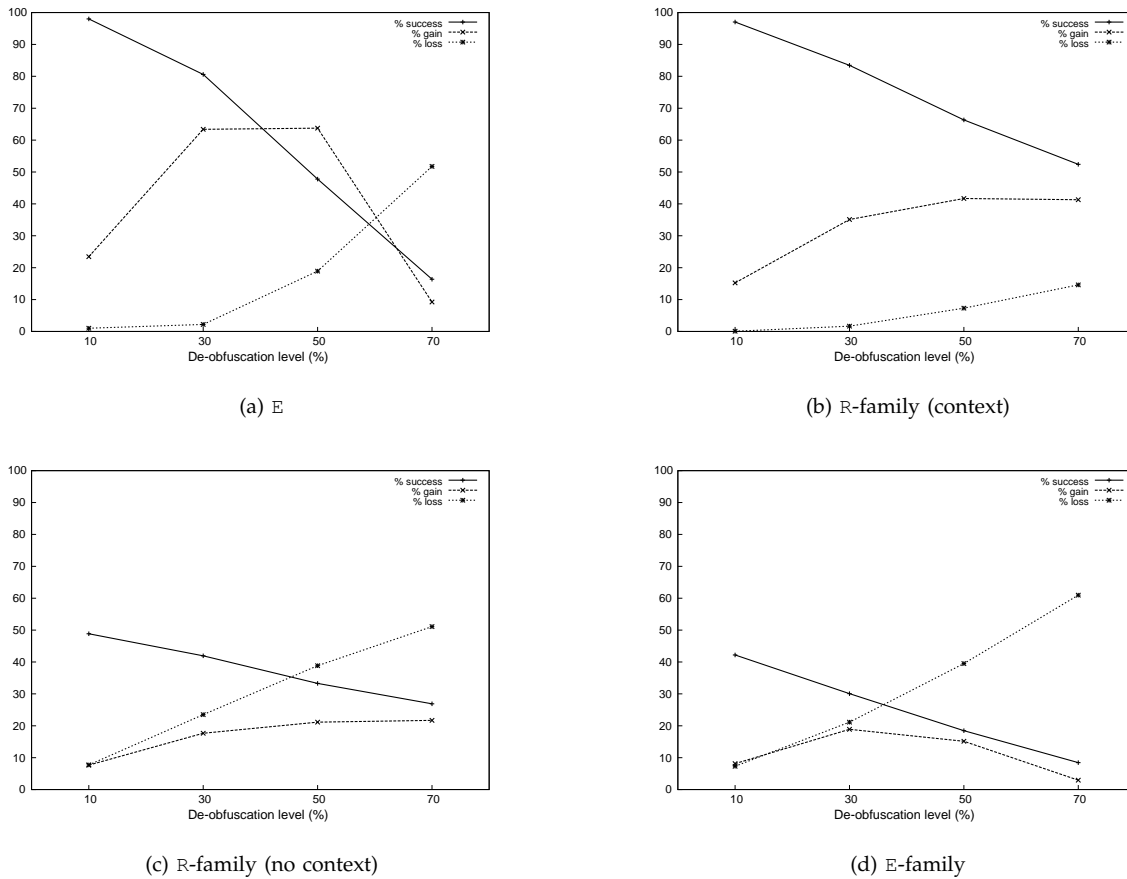


Fig. 14. Adversary successes, gains, and losses

- in the R-family with contextual awareness, the adversary always succeeds except for very high de-obfuscation levels. This behavior is due to operator S that, when used in the R-family, reduces the probability of exceeding the breakpoint and, therefore, of retrieving a resulting de-obfuscation relevance less than the initial one;
- operator E performs adequately only for high de-obfuscation levels (more than 50%), since, in average, when the radius is considerably enlarged the de-obfuscation fails.

Analyzing just the success rates, we can conclude that:

- the E-family is the most robust, since it exhibits the lowest success rates among all operators;
- the R-family obfuscation is highly sensitive to the adversary's contextual awareness: if the adversary cannot infer the type of obfuscation it provides strong obfuscation; otherwise the resulting obfuscation is weak;
- operator E is robust against greedy adversaries only. Most conservative adversaries, which de-obfuscate up to 50%, mostly succeed.

Adversary gain analysis. We have tested how relevance gains and losses vary by increasing the levels of de-obfuscation. When the strategy adopted is suitable to de-

obfuscate the obfuscation operator under consideration, the risk for the adversary is to exceed in the radius modification and produce a de-obfuscated area with a relevance less than the one of the obfuscated area. This analysis is useful because a rational adversary will look for that amount of radius manipulation that maximizes the combination of the success rate and relevance gain achieved, while minimizing the relevance loss. Formally, we define the *utility function* for the adversary as $U = W \cdot G - (1 - W) \cdot L$, where W is the rate of success, G is the mean gain, and L is the mean loss. U assumes values in $[-1, 1]$ where positive values represent an incentive to de-obfuscate; negative values indicate a disincentive to de-obfuscate; and $U = 0$ represents neutrality.

For each adversary behaviour, Figure 14 illustrates the success rate, and the mean relevance gain and loss. Mean gain and loss are obtained by calculating the value returned by each de-obfuscation attempt for every obfuscated area in our dataset, and then by computing the mean for each de-obfuscation level. Figure 15 compares the utility function values (y-axis) for the different rate of radius modification (x-axis).

Operator E. Figure 14(a) shows the results for operator E. The maximum mean gain is obtained for radius manipulations of 50%, corresponding to a relevance gain of 64%. The mean loss is lower than 20% up to 50% of radius

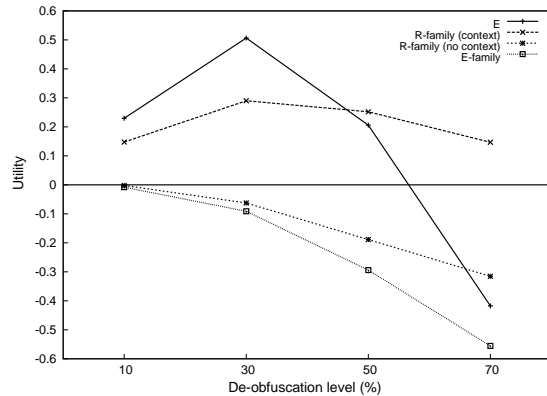


Fig. 15. Adversary's utility function

manipulation and increases for larger de-obfuscation levels. The adversary utility function, as shown in Figure 15, is maximum (with a value of 0.5) in correspondence of 30% of radius manipulation.

R-family. For the *R-family*, we have to consider two scenarios. Figure 14(b) shows the case of a contextual aware adversary. The maximum mean gain is 42%, achieved for 50% of radius enlargement. The adversary utility function (Figure 15) is maximum (with a value of 0.29) for 30% of radius manipulation. Figure 14(c) shows the case of an adversary with no contextual information. Here, the maximum gain is 21%, achieved at both 50% and 70% of radius manipulation, and the utility function returns an increasingly negative value for every radius manipulation. A de-obfuscation level of 10% gives a value of -0.002.

E-family. Figure 14(d) shows the results produced when the *E-family* operators are de-obfuscated. The highest relevance gain is 18% corresponding to a radius manipulation of 30%. The utility function has the same shape as for the *R-family* with no contextual awareness and is always negative. A de-obfuscation level of 10% gives a value of -0.008.

Analyzing these results, we can conclude that:

- only when a preferred strategy exists (i.e., operator *E* only and the *R-family* with contextual awareness) the adversary has an incentive to de-obfuscate and 30% of radius manipulation is the best choice;
- when the adversary acts randomly (i.e., the *R-family* without contextual awareness and the *E-family*) there is no incentive to de-obfuscate because for every radius manipulation the utility function returns negative values;
- operator *E* is the weakest obfuscation operator because, regardless to any contextual information known by the adversary, it provides the highest utility function value among the families;
- the *E-family* is the strongest obfuscation family because, regardless to any contextual information known by the adversary, the success rate is always

less than 50% and the adversary's utility function is always negative;

- the robustness of the *R-family* depends on the adversary awareness. When the adversary knows that the *R-family* has been used for obfuscation, the *R-family* is weak and exhibits a behavior similar to operator *E*. Instead, when the adversary has no contextual information, the *R-family* is strong and similar to the *E-family*.

8 A MOBILE SOCIAL NETWORK SCENARIO

A Mobile Social Network (MSN) represents a suitable application scenario for our obfuscation techniques since it can be easily enriched with location information. For instance, Loopt [26] is an available online service that locates friends through cell phones by georeferentiating GPS coordinates on a map. Each Loopt user can decide to share or not the information about her physical position on a friend-by-friend basis or for all friends at once. Users of MSN cannot be anonymized, are often involved in large web of relation (friends, co-workers, relatives, or just acquaintances), and typically restrict information made available to others according to the type of relationship or on a person-by-person basis. Location information could be managed in a similar way by integrating our obfuscation-based solution into the MSN. A typical scenario may involve a user that requires the position of a person in her own web of relation or asks the MSN for users in her proximity. Such a location information should be managed according to the privacy preferences of all users involved. A possible architecture could include a *trusted middleware*, implementing our techniques, which receives location requests and privacy preferences from the MSN, retrieves actual locations from a location service (e.g., a cell phone operator), and returns them to the MSN, obfuscated according to the privacy preferences of the users.

The adversary could be any user of the MSN who may want to breach the location privacy required by a person in her web of relation. With our solution, a potential adversary receives an obfuscated location and can just manipulate it trying to achieve a more accurate area.

The MSN management system can be considered a trusted party since it manages users privacy preferences (i.e., different mobile services would manage their own users location privacy preferences). In a different setup, we could imagine that mobile services are untrusted, therefore the middleware should centralize and manage users privacy preferences and apply them to all location requests. In either cases, the architecture does not affect the application of our obfuscation operators.

9 RELATED WORK

Location-based information and its management have been considered in several works in the area of mobile applications, including approaches aimed at protecting the privacy of users. Some works are based on the

definition of *privacy policies* that define restrictions that must be enforced when the location information is used by or released to external parties [17], [24].

The line of research closest to the work in this paper exploits obfuscation as the process of degrading the accuracy of the location information to provide privacy protection. Obfuscation-based techniques perturb the location information while maintain a binding with the users identities. Duckham and Kulik [14] present a framework that provides a mechanism for balancing the individuals needs for high-quality information services and the location privacy. The authors propose to degrade the quality of the location information and to provide obfuscation features by adding n points at the same probability to the real user position. In general, all these obfuscation solutions share some common drawbacks. First, they do not provide a quantitative estimation of the provided privacy level, making them difficult to integrate into a full fledged location-based application scenario [1]. Second, such solutions implement a single obfuscation technique based on the enlargement of the location area whose effect can be easily undid by the adversary. Our previous works [2], [3] address these shortcomings by presenting some techniques aimed at preserving location privacy by artificially perturbing location information. In this paper, we substantially improve our previous proposals by first providing the probabilistic fundamentals of our obfuscation operators, by showing how these operators can be composed, by evaluating the robustness of the operators against de-obfuscation attempts performed by adversaries, and by showing some experiments that validate our solution.

Another important line of research exploits the concept of anonymity to provide techniques suitable when the identity of the users is not relevant for the provision of a service. Beresford and Stajano [6], [7] introduce a solution based on the concepts of *application zones*, representing similar application interests in specific geographic areas, and *mix zones*, which are areas where a user cannot be tracked. Within each mix zone, the identities of all users are mixed and become indiscernible, and users entering the mix zone are unlinkable from other users leaving it. Bettini et al. [8] propose a framework for evaluating the risk of disseminating sensitive location-based information, and introduce a technique aimed at supporting k -anonymity [10], [30]. The authors put forward the idea that the geo-localized history of the requests submitted by a user can be considered as a *quasi-identifier*, that is, a set of attributes that can be linked with external information, thus reducing the uncertainty over the identity of the user. The service provider gathering both users requests and personal histories of locations should be able to link a request to at least $k-1$ users having a personal history of locations compatible with the issued requests. Gruteser and Grunwald [19] propose a middleware architecture and an algorithm to adjust location information resolution, in spatial or temporal dimensions, to comply with a

specific k -anonymity requirement. Gedik and Liu [16] describe another k -anonymity model where each user is able to define the minimum level of anonymity and the maximum acceptable temporal and spatial resolution for her location measurement. They define a message perturbation engine responsible for providing location anonymization of user's requests through identity removal and spatio-temporal obfuscation of location information. Mokbel et al. [27] present a framework where each user defines her privacy preferences through a parameter k , which is the k -anonymity requirement of the user, and an area A_{min} that is the minimum acceptable resolution of her location information. That framework includes a *location anonymizer*, for perturbing the location information of users to achieve their privacy preferences, and a *privacy-aware query processor*, for the management of anonymous queries and cloaked spatial areas. Ghinita et al. [18] propose PRIVÉ, a decentralized architecture for preserving query anonymization based on the definition of k -anonymous areas. A common drawback of all these anonymity-based techniques is that their applicability and performances depend on the number of users physically located in a particular area. Another line of research, which is not directly related to our work, is aimed at protecting the path privacy of the users [21], [22].

10 CONCLUSIONS

We presented different obfuscation operators that protect the location privacy of users by changing their location information. Our proposal takes into consideration both the accuracy of location measurements and the users needs of privacy. We also provided an evaluation of the robustness of such operators. The analysis and the experimental results prove that our operators provide better protection than the simple enlargement usually applied by current solutions.

The work presented in this paper leaves space for further work: the analysis of our solution assuming Gaussian-like distributions and complex location measurement shapes; the introduction of map constraints in the computation of obfuscated areas; the definition of additional techniques for degrading the temporal accuracy of location measurements; the extension of our solution to protect the path privacy of the users; and the actual integration and extensive test of our solution in a real scenario.

ACKNOWLEDGMENTS

This work was supported in part by the EU, within the 7th Framework Programme (FP7/2007-2013) under grant agreement no. 216483 "PrimeLife".

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 ACM ASIACCS 2006*, Taipei, Taiwan, March 2006.

- [2] C.A. Ardagna, M. Cremonini, E. Damiani, S. De Capitani di Vimercati, and P. Samarati. A middleware architecture for integrating privacy preferences and location accuracy. In *Proc. of IFIP SEC 2007*, Sandton, South Africa, May 2007.
- [3] C.A. Ardagna, M. Cremonini, E. Damiani, S. De Capitani di Vimercati, and S. Samarati. Location privacy protection through obfuscation-based techniques. In *Proc. of IFIP DBSEC 2007*, Redondo Beach, CA, USA, July 2007.
- [4] L. Barkhuus and A. Dey. Location-based services for mobile telephony: A study of user's privacy concerns. In *Proc. of IFIP INTERACT 2003*, 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 ISWCS 2005*, Siena, Italy, September 2005.
- [6] A.R. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.
- [7] A.R. Beresford and F. Stajano. Mix zones: User privacy in location-aware services. In *Proc. of IEEE PERCOMW 2004*, Orlando, FL, USA, March 2004.
- [8] 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*, Trondheim, Norway, 2005.
- [9] Chicago Tribune. *Rental firm uses GPS in speeding fine*. July 2nd, p9. Associated Press: Chicago, IL, 2001.
- [10] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati. k-Anonymity. In T. Yu and S. Jajodia, editors, *Secure Data Management in Decentralized Systems*. Springer-Verlag, 2007.
- [11] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati. Microdata protection. In T. Yu and S. Jajodia, editors, *Secure Data Management in Decentralized Systems*. Springer-Verlag, 2007.
- [12] E. Damiani, M. Anisetti, and V. Bellandi. Toward exploiting location-based and video information in negotiated access control policies. In *Proc. of ICISS 2005*, Kolkata, India, December 2005.
- [13] T. D'Roza and G. Bilchev. An overview of location-based services. *BT Technology Journal*, 21(1):20–27, January 2003.
- [14] M. Duckham and L. Kulik. A formal model of obfuscation and negotiation for location privacy. In *Proc. of PERVASIVE 2005*, Munich, Germany, May 2005.
- [15] M. Duckham and L. Kulik. Dynamic & mobile GIS: Investigating change in space and time. In *Location privacy and location-aware computing*. Taylor & Francis, 2006.
- [16] B. Gedik and L. Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*, 7(1):1–18, January 2008.
- [17] Geographic Location/Privacy (geopriv), September 2006. <http://www.ietf.org/html.charters/geopriv-charter.html>.
- [18] G. Ghinita, P. Kalnis, and S. Skiadopoulos. Privé: Anonymous location-based queries in distributed mobile systems. In *Proc. of WWW 2007*, Banff, Canada, May 2007.
- [19] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proc. of MobiSys 2003*, San Francisco, CA, USA, May 2003.
- [20] F. Gustafsson and F. Gunnarsson. Mobile positioning using wireless networks: Possibilities and fundamental limitations based on available wireless network measurements. *IEEE Signal Processing Magazine*, July 2005.
- [21] B. Ho and M. Gruteser. Protecting location privacy through path confusion. In *Proc. of IEEE/CreateNet SecureComm 2005*, Athens, Greece, September 2005.
- [22] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Preserving privacy in GPS traces via density-aware path cloaking. In *Proc. of ACM CCS 2007*, Alexandria, VA, USA, October 2007.
- [23] M. Langheinrich. Privacy by design-principles of privacy-aware ubiquitous systems. In *Proc. of UBIComp 2001*, Atlanta, Georgia, USA, September-October 2001.
- [24] M. Langheinrich. A privacy awareness system for ubiquitous computing environments. In *Proc. of UBIComp 2002*, Goteborg, Sweden, September-October 2002.
- [25] J-W. Lee. *Location-tracing sparks privacy concerns*. Korea Times. <http://times.hankooki.com>, 16 November 2004. Accessed 22 December 2006.
- [26] Loopt. <http://www.loopt.com/>, December 2008.
- [27] M.F. Mokbel, C-Y. Chow, and W.G. Aref. The new casper: Query processing for location services without compromising privacy. In *Proc. of VLDB 2006*, Seoul, Korea, September 2006.
- [28] P. Olofsson. *Probability, Statistics and Stochastic Processes*. John Wiley & Sons, Inc., 2005.
- [29] Privacy Rights Clearinghouse/UCAN. *A Chronology of Data Breaches*, 2006. <http://www.privacyrights.org/ar/ChronDataBreaches.htm>.
- [30] P. Samarati. Protecting respondents' identities in microdata release. *IEEE TKDE*, 13(6):1010–1027, November/December 2001.
- [31] H. Shin, V. Atluri, and J. Vaidya. A profile anonymization model for privacy in a personalized location based service environment. In *Proc. of MDM 2008*, Beijing, China, April 2008.
- [32] 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*, July 2005.
- [33] B. Thuraisingham. Dependable infrastructures and data managers for sensor networks. In *Proc. of IEEE WORDS 2003*, Guadalajara, Mexico, October 2003.
- [34] B. Thuraisingham. Directions for security and privacy for semantic e-business applications. *Communications of the ACM (CACM)*, 48(12):71–73, December 2005.
- [35] B. Thuraisingham. Privacy constraint processing in a privacy-enhanced database management system. *Data & Knowledge Engineering*, 55(2):159–188, November 2005.



Claudio A. Ardagna is an assistant professor at the Department of Information Technology, Università degli Studi di Milano, Italy. He received the laurea and PhD degrees, both in computer science, from the Università degli Studi di Milano in 2003 and 2008, respectively. His research interests are in the area of information security, privacy, access control, mobile networks, and open source. <http://www.dti.unimi.it/ardagna>



Marco Cremonini is an assistant professor at the Department of Information Technology of the University of Milan, Italy. He previously worked as a Research Assistant at the Institute for Security Technology Studies (ISTS) of the Dartmouth College, NH, USA. His research activity is focused on network security, economic aspects of security technologies, privacy, and security in ubiquitous computing.



Sabrina De Capitani di Vimercati is a professor at the Department of Information Technology, Università degli Studi di Milano, Italy. Her research interests are in the area of information security, databases, and information systems. She has been an international fellow in the Computer Science Laboratory at SRI, CA (USA). She is co-recipient of the ACM-PODS'99 Best Newcomer Paper Award. <http://www.dti.unimi.it/decapita>



Pierangela Samarati is a professor at the Department of Information Technology, Università degli Studi di Milano, Italy. Her main research interests are in data protection, access control models, and information privacy and security. She has published more than 150 papers in international journals and conferences. She has been a computer scientist at SRI International, CA (USA) and a visiting researcher at Stanford University, CA (USA), and George Mason University, VA (USA). She is the chair of the

Steering Committees of the ACM Workshop on Security and Privacy, and of the European Symposium on Research in Computer Security. She is a member of the steering committee of several conferences. She is the vice-chair of the ACM SIGSAC - Special Interest Group on Security, Audit, and Control. <http://www.dti.unimi.it/samarati>