

# An Obfuscation-Based Approach for Protecting Location Privacy

Claudio A. Ardagna, Marco Cremonini, Sabrina De Capitani di Vimercati, and 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 (on the contrary, GPS is operating mainly outdoor). Key aspects of our perturbation process, called *obfuscation*, are: 1) to allow users to express their privacy preferences in a simple and intuitive way and 2) to enforce the privacy preferences through a set

• The authors are with the Dipartimento di Tecnologie dell'Informazione (DTI), Università degli Studi di Milano, 65, Crema (CR) 26013, Italy. E-mail: {claudio.ardagna, marco.cremonini, sabrina.decapitani, pierangela.samarati}@unimi.it.

Manuscript Received 5 July 2008; revised 14 Dec. 2008; accepted 26 May 2009; published online 11 June 2009.

For information on obtaining reprints of this article, please send e-mail to: [tdsc@computer.org](mailto:tdsc@computer.org), and reference IEEECS Log Number TDSC-2008-07-0101. Digital Object Identifier no. 10.1109/TDSC.2009.25.

of techniques robust against a relevant class of deobfuscation 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.<sup>1</sup> 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 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  $\delta r$  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 a dimensional 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 as  $\mathcal{R}_i$ , 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; and
- 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

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.

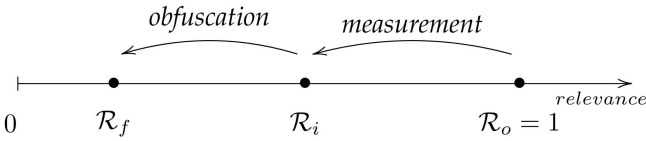


Fig. 1. Relevance degradation due to the intrinsic measurement error and obfuscation.

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 as  $\lambda$ , 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 of basic *obfuscation operators* (or a combination of them) that change the radius, or the center, of the original location measurement. As illustrated in Fig. 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 as  $f_i(X, Y)$ , 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  $\lambda$  (see Definition 2.4) that can be defined as the composite probability of the following two independent events: 1) a random point  $(x', y') \in A_f$  belongs to the intersection between  $A_i$  and  $A_f$  and 2) the user's actual position  $(x_u, y_u)$  belongs to the intersection. The term  $\lambda$  is then equal to

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

From Definition 2.4 and (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  $\lambda$  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 Fig. 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  $\pi r_i^2$  in Fig. 2, is equal to 1. It is easy to say that the random variable  $C$  is statistically equivalent to variables  $(X, Y)$ .

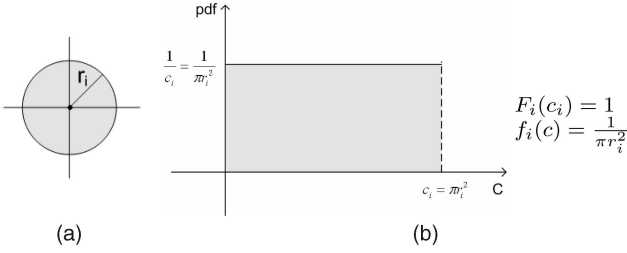


Fig. 2. (a) A location measurement and (b) the pdf of the corresponding variable  $C$ .

#### 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  $op: \mathcal{A} \times (0, 1] \times (0, 1] \rightarrow \mathcal{A}$  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.

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  $E(A_i, \mathcal{R}_i, \mathcal{R}_f) = A_f$  with radius  $r_f > r_i$  (see Fig. 3a). 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  $C$ , Fig. 3a 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 (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

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

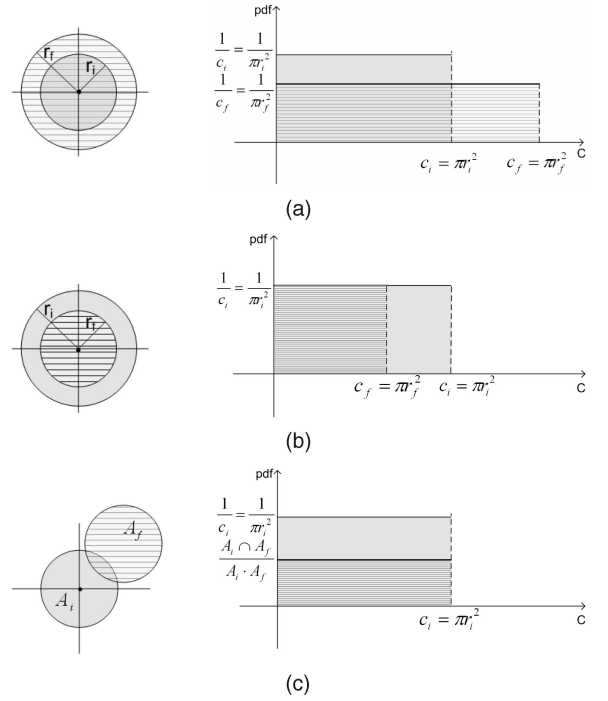


Fig. 3. Graphical illustration of the basic obfuscation operators and their probabilistic effects on variable  $C$ . (a) Radius enlargement. (b) Radius reduction. (c) Center shifting.

For instance, suppose that the privacy preference of the user is  $\mathcal{R}_f = 0.16$  and a location measurement with the 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 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  $R(A_i, \mathcal{R}_i, \mathcal{R}_f) = A_f$  with radius  $r_f < r_i$  (see Fig. 3b). 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.,  $f_f(c) = f_i(c) = \frac{1}{\pi r_i^2}$ ). Considering variable  $C$ , Fig. 3b 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.,  $F_f(c_f) < F_i(c_i) = 1$ ).

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  $\mathcal{R}_f$ :

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.

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

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

For instance, suppose that the privacy preference of the user is  $\mathcal{R}_f = 0.16$  and a location measurement with the 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 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  $S(A_i, \mathcal{R}_i, \mathcal{R}_f) = A_f$  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  $A_f$ , and  $r_f = r_i$  (see Fig. 3c). 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  $C$ , Fig. 3c 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  $\theta$  must be specified to derive an obfuscated area by shifting the center. For the scope of this paper, and without loss of generality,  $\theta$  is assumed to be randomly generated. Strategies for selecting a value for  $\theta$  depend on the application context [2]. From (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:

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

By 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  $\sigma$  and  $\gamma$  are the central angles of the circular sectors identified by the two radii connecting the center of  $A_i$  and  $A_f$  with the intersection points of  $A_i$  and  $A_f$ , 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

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

Fig. 4. Basic obfuscation operators.

(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  $\theta$ , 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 a location measurement with the 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 (9), the application of operator 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  $\theta$  is selected and the obfuscated area is generated.

Fig. 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.

## 5 COMPOSITION OF THE BASIC OBFUSCATION OPERATORS

The basic obfuscation operators just illustrated transform a location measurement by changing its radius (operators E and R) or by changing its center (operator 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

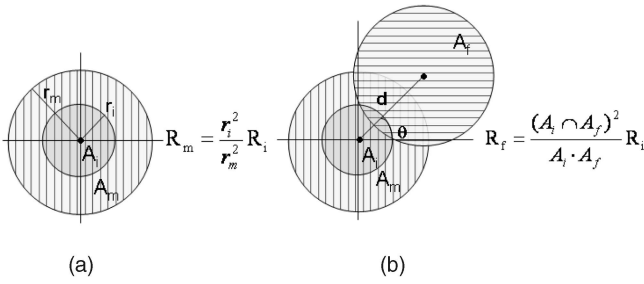


Fig. 5. Composed operator ES applied on  $A_i$ . (a) First obfuscation step. (b) Second obfuscation step.

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  $\text{op}(A_i, \mathcal{R}_i, \mathcal{R}_f) \equiv \text{op}_{A_i, \mathcal{R}_i}(A_i, \mathcal{R}_f)$ , 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 as  $hk$  (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 the composed operator  $\text{ES} = \text{S}_{A_i, \mathcal{R}_i}(\text{E}_{A_i, \mathcal{R}_i}(A_i, \mathcal{R}_m), \mathcal{R}_f)$  illustrated in Fig. 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), and  $A_f$  is the obfuscated area produced by the second operator (the area filled with horizontal lines). In the first obfuscation step (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  $r_m = r_i \sqrt{\frac{\mathcal{R}_i}{\mathcal{R}_m}}$  (see (4)). According 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 E, R, and 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 E and R with operator 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 Fig. 6).

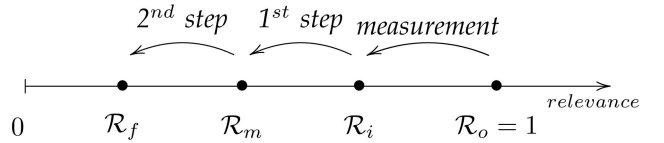


Fig. 6. Relevance degradation due to intrinsic measurement error and obfuscation.

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

Fig. 7. Redefinition of the obfuscation operators.

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. Fig. 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: 1) in the first step (1st) of a composed operator to produce, starting from the original location measurement  $A_i$ , an intermediate area with relevance  $\mathcal{R}_m$  and 2) 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 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 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 shown 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.  $\mathcal{A}_{A_i, \mathcal{R}_i}^{\text{SE}} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{ES}};$
2.  $\mathcal{A}_{A_i, \mathcal{R}_i}^{\text{SR}} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{RS}};$
3.  $\mathcal{A}_{A_i, \mathcal{R}_i}^{\text{ES}} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{SE}};$
4.  $\mathcal{A}_{A_i, \mathcal{R}_i}^{\text{RS}} \subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{SR}}.$

**Proof.**

1.  $\mathcal{A}_{A_i, \mathcal{R}_i}^{\text{SE}} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{ES}}.$  Let  $A_f \in \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{SE}}$  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 (see (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.  $\mathcal{A}_{A_i, \mathcal{R}_i}^{\text{SR}} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{RS}}.$  Let  $A_f \in \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{SR}}$  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 could never be returned by the second (shifting) step.
3.  $\mathcal{A}_{A_i, \mathcal{R}_i}^{\text{ES}} \not\subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{SE}}.$  Let  $A_f \in \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{ES}}$  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.  $\mathcal{A}_{A_i, \mathcal{R}_i}^{\text{RS}} \subseteq \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{SR}}.$  It is easy to see that  $\forall A_f \in \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{RS}}$  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  $\forall A_f \in \mathcal{A}_{A_i, \mathcal{R}_i}^{\text{RS}}$ ,  $A_f$  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.  $\square$

Fig. 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  $A_i$  and 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.

## 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 deobfuscation 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 deobfuscated 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 deobfuscation attempt.

While it is relatively straightforward to deobfuscate an area by applying some transformations, understanding whether the deobfuscated area is more or less accurate than the obfuscated area could be an irresolvable task for the adversary. In this situation, called *blind deobfuscation*, 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 deobfuscation 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: 1) the *obfuscated area*; 2) the *location sensing technology* adopted by the location service; and 3) *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

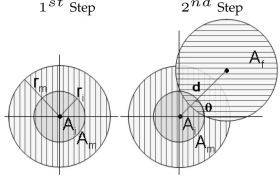
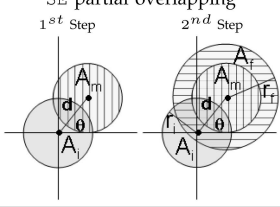
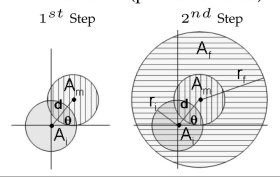
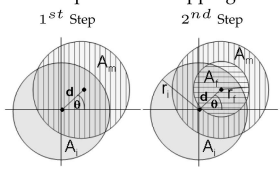
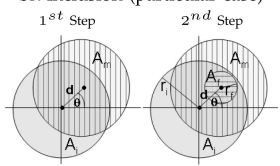
Op	Schema	1 <sup>st</sup> step ( $A_m = \langle x_m, y_m, r_m \rangle$ )			2 <sup>nd</sup> step ( $A_f = \langle x_f, y_f, r_f \rangle$ )			Comment
		$x_m$	$y_m$	$r_m$	$x_f$	$y_f$	$r_f$	
ES		$x_i$	$y_i$	$r_i \sqrt{\frac{\mathcal{R}_i}{\mathcal{R}_f}}$	$x_m + d \sin \theta$	$y_m + d \cos \theta$	$r_m$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

Fig. 8. Composed operators.

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 possible 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 knows only 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  $\theta$ ) that the adversary cannot evaluate. The consequence is that if the adversary tries to deobfuscate the given obfuscated area through a shifting of the center, it cannot evaluate whether the deobfuscated area has a relevance greater than the relevance associated with the obfuscated area. Therefore, we assume that the adversary tries to deobfuscate 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 deobfuscation 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



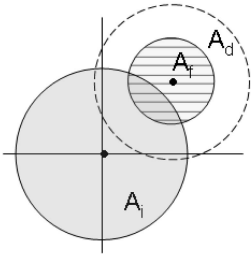


Fig. 9. Deobfuscation attempt on area  $A_f$  produced through composed operator SR (partial overlapping).

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.

### 6.1 R-Family Deobfuscation

The R-family deobfuscation 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 Fig. 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 deobfuscated 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 (7)). For each operator of the R-family, Fig. 10 shows the variation of the relevance (Y-axis) as a function of the radius of the deobfuscated area (X-axis), where the result of a deobfuscation attempt is intuitively represented by the + and – labels: a deobfuscation 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 deobfuscation attempt;
- radius  $r_{max}$  of the area with the best relevance  $\mathcal{R}_{max}$ , which represents the best deobfuscation that the adversary can achieve;
- radius  $r_{i,d}$  of the deobfuscated 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  $A_f$ ; and
- 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 deobfuscated area of less relevance than  $\mathcal{R}_f$ .

From Fig. 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

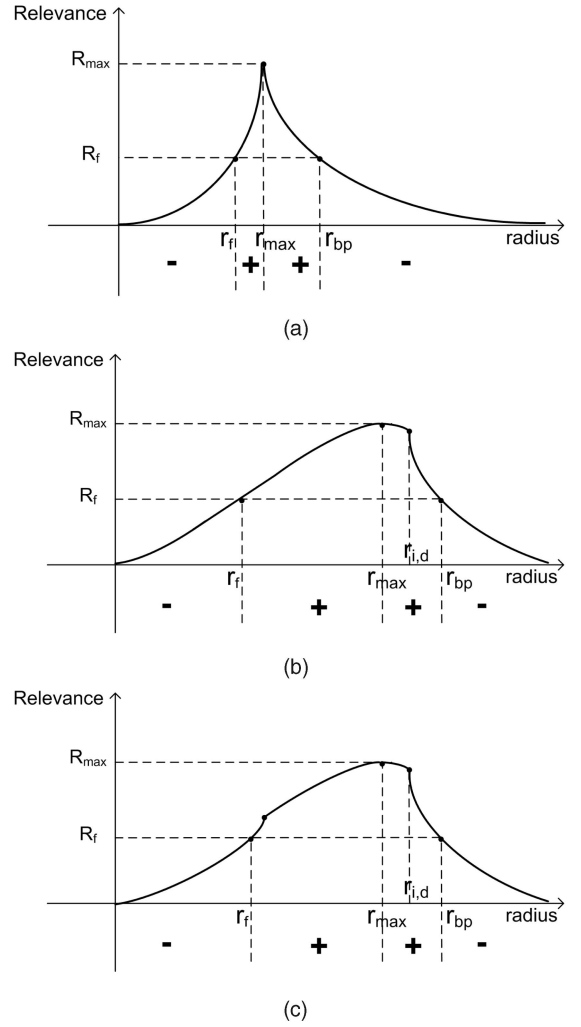


Fig. 10. Relevance variations in deobfuscation attempts against the R-family. (a) R. (b) SR (partial overlapping). (c) SR (inclusion).

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. 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 Fig. 10a) are based on the quadratic function of operator R (see (5)), between  $r_f$  and  $r_{max}$ , and on the quadratic function of operator E (see (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 Fig. 10b) are based on the partial overlapping produced by operator S (see (7)),

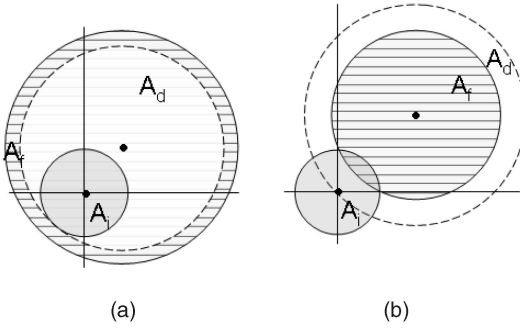


Fig. 11. Deobfuscation attempt on area  $A_f$  produced via (a) operator SE (inclusion) and (b) operator ES.

between  $r_f$  and  $r_{i,d}$ , and the quadratic function of operator E (see (3)), between  $r_{i,d}$  and  $r_{bp}$ . In this 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 deobfuscated 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 Fig. 10c) follows the quadratic function of operator R (see (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 deobfuscation 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 Deobfuscation

Although it may seem that to deobfuscate 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 deobfuscated 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 Fig. 11, where  $A_i$  is the original location measurement,  $A_f$  is the obfuscated area, and  $A_d$  is the deobfuscated area produced by manipulating the radius of area  $A_f$ . Fig. 11a 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 (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. 11b illustrates instead an obfuscated area  $A_f$  obtained through operator ES and a deobfuscated area  $A_d$ , with a relevance better than the

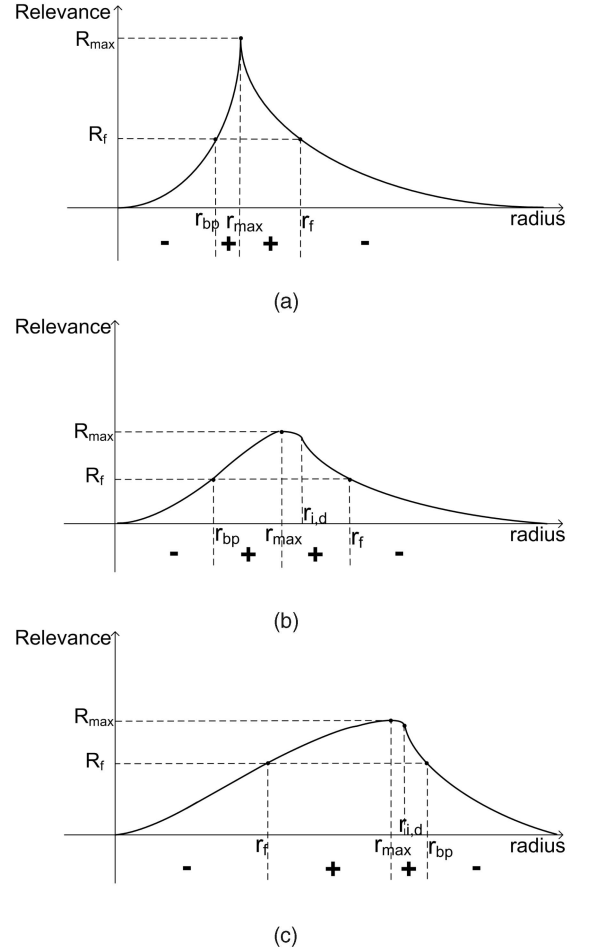


Fig. 12. Relevance variations in deobfuscation attempts against the E-family. (a) E. (b) SE (inclusion). (c) ES/SE (partial overlapping).

relevance associated with  $A_f$ , obtained by enlarging, rather than reducing, the radius of the obfuscated area. Considering (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, Fig. 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 deobfuscated area with maximal relevance, the radius of the deobfuscated 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) (Figs. 12a and 12b):** 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. Fig. 12a shows that for operator E, the relevance variation obtained by reducing the radius  $r_f$  is modeled by the quadratic function of

operator E (see (3)), between  $r_{max}$  and  $r_f$ , and is modeled as an obfuscation produced by reducing the radius (see (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  $\mathcal{R}_{max} = \mathcal{R}_i$ . Fig. 12b 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 (see (3)), between  $r_{i,d}$  and  $r_f$ , and is then modeled as a function of the overlapping of the areas (see (7)), between  $r_{bp}$  and  $r_{i,d}$ . A maximum relevance  $\mathcal{R}_{max} < \mathcal{R}_i$  can be achieved by the adversary with radius  $r_{max}$ .

**Case 2. Operators ES and SE (partial overlapping) (Fig. 12c):** 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. Fig. 12c 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 (see (7)), between  $r_f$  and  $r_{i,d}$ , and by the quadratic function of operator E (see (3)), between  $r_{i,d}$  and  $r_{bp}$ .

From our analysis, we can conclude that for the E-family, there is not a preferred deobfuscation 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 deobfuscated area with respect to the obfuscated one, the E-family is said to be strongly robust.

### 6.3 \*-Family Deobfuscation

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 deobfuscated by enlarging its radius, while an obfuscated area produced through obfuscation operators SE (inclusion) and E should be deobfuscated by reducing its radius. The radius enlargement is then the most likely deobfuscation 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 data set 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 data sets 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 data sets of 20,000 obfuscated areas each produced by applying: 1) the operators belonging to the R-family, randomly; 2) the operators belonging to the E-family, randomly; and 3) operator 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 deobfuscation strategies. We considered two main adversary behaviors: 1) *no contextual awareness*, when the adversary is not aware of any contextual information and is not able to infer the obfuscation family applied and 2) *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 behavior that has to be assumed during the evaluation of the R-family: 1) the adversary with no contextual information does not know that an operator of the 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 and 2) the adversary that knows that one operator of the R-family has been applied will only enlarge the obfuscated area.

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

Finally, we assume that deobfuscation attempts consist in enlarging/reducing the radius of the obfuscated area by a deobfuscation level of 10, 30, 50, and 70 percent. This is aimed to test different adversary behaviors, from the most conservative to the greediest. The hypothesis is that high deobfuscation levels are associated with both high gains in relevance and low success rates, while low deobfuscation levels result in low gains and high 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 deobfuscation rate achieved by the adversary and the relevance gain or loss the adversary obtains as a result of a deobfuscation 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 deobfuscated area has a relevance greater than the one associated with the obfuscated area. The data set produced by applying the operators of the R-family has been tested twice for the two different adversary behaviors depending on the presence or absence of contextual awareness.

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

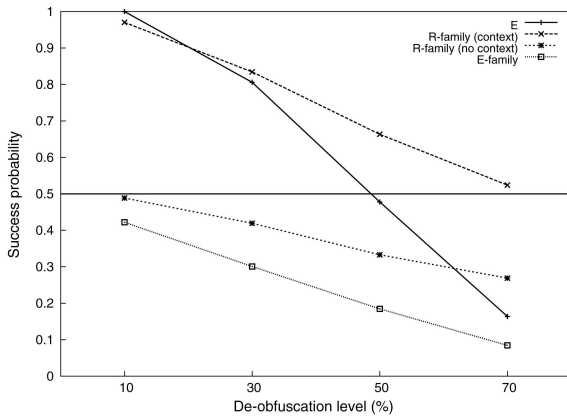


Fig. 13. Rate of successful deobfuscations attempts based on different degrees of manipulations.

- deobfuscation attacks against the E-family and the R-family with no contextual awareness never exceed a success rate of 50 percent, 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 deobfuscation succeeds;
- deobfuscation attacks against operator E and the R-family with contextual awareness have a success rate ranging from more than 95 percent for very small radius modification to 80 percent for a deobfuscation of 30 percent;
- in the R-family with contextual awareness, the adversary always succeeds except for very high deobfuscation 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 deobfuscation relevance less than the initial one;
- operator E performs adequately only for high deobfuscation levels (more than 50 percent), since, on average, when the radius is considerably enlarged, the deobfuscation 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; and
- operator E is robust against greedy adversaries only. Most conservative adversaries, which deobfuscate up to 50 percent, mostly succeed.

**Adversary gain analysis.** We have tested how relevance gains and losses vary by increasing the levels of deobfuscation. When the strategy adopted is suitable to deobfuscate the obfuscation operator under consideration, the risk for the adversary is to exceed in the radius modification and produce a deobfuscated 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 the values in  $[-1, 1]$ , where positive values represent an incentive to deobfuscate; negative values indicate a disincentive to deobfuscate; and  $U = 0$  represents neutrality.

For each adversary behavior, Fig. 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 deobfuscation attempt for every obfuscated area in our data set, and then, by computing the mean for each deobfuscation level. Fig. 15 compares the utility function values (y-axis) for the different rate of radius modification (x-axis).

**Operator E.** Fig. 14a shows the results for operator E. The maximum mean gain is obtained for radius manipulations of 50 percent, corresponding to a relevance gain of 64 percent. The mean loss is lower than 20 up to 50 percent of radius manipulation and increases for larger deobfuscation levels. The adversary utility function, as shown in Fig. 15, is maximum (with a value of 0.5) in correspondence of 30 percent of radius manipulation.

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

**E-family.** Fig. 14d shows the results produced when the E-family operators are deobfuscated. The highest relevance gain is 18 percent corresponding to a radius manipulation of 30 percent. The utility function has the same shape as for the R-family with no contextual awareness and is always negative. A deobfuscation level of 10 percent 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 deobfuscate and 30 percent 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 deobfuscate 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 percent and the adversary's utility function is always negative; and

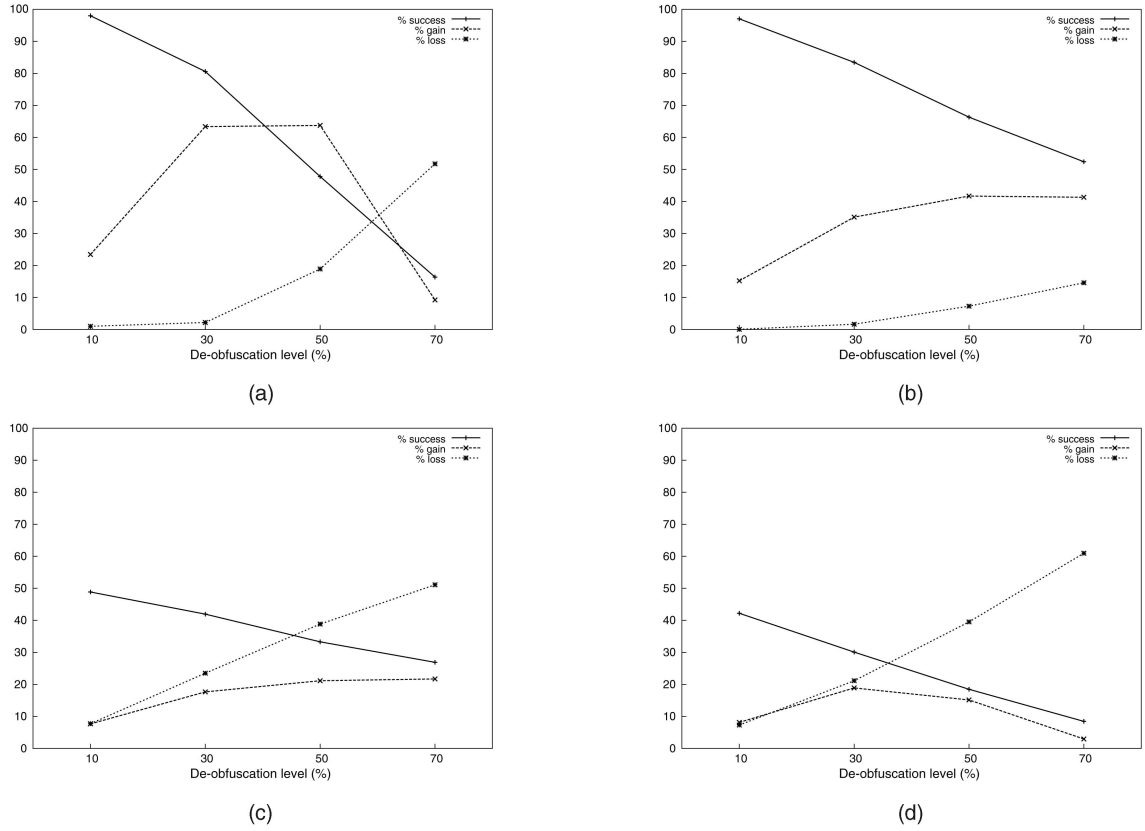


Fig. 14. Adversary successes, gains, and losses. (a) E. (b) R-family (context). (c) R-family (no context). (d) E-family.

- 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, coworkers, 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.

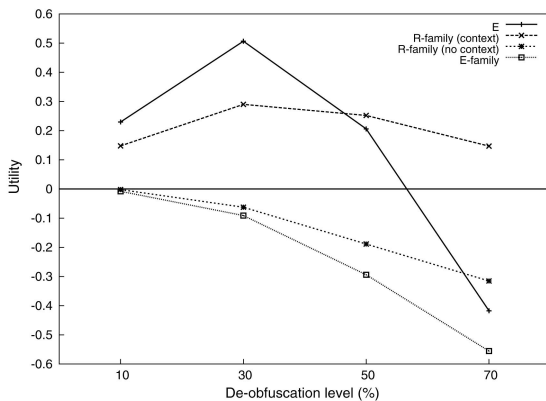


Fig. 15. Adversary's utility function.

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 maintaining 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 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 reversed 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 deobfuscation 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 the 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 geolocalized 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 spatiotemporal 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 PRIVÈ, 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 European Union, within the seventh 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," *Proc. ACM Symp. Information, Computer and Comm. Security (ASIACCS '06)*, Mar. 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," *Proc. IFIP Int'l Information Security Conf. (SEC '07)*, 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," *Proc. IFIP Working Conf. Data and Applications Security (DBSEC '07)*, July 2007.
- [4] L. Barkhuus and A. Dey, "Location-Based Services for Mobile Telephony: A Study of User's Privacy Concerns," *Proc. IFIP Int'l Conf. Human-Computer Interaction (INTERACT '03)*, Sept. 2003.
- [5] P. Bellavista, A. Corradi, and C. Giannelli, "Efficiently Managing Location Information with Privacy Requirements in Wi-Fi Networks: A Middleware Approach," *Proc. Int'l Symp. Wireless Comm. Systems (ISWCS '05)*, Sept. 2005.
- [6] A.R. Beresford and F. Stajano, "Location Privacy in Pervasive Computing," *IEEE Pervasive Computing*, vol. 2, no. 1, pp. 46-55, Jan.-Mar. 2003.
- [7] A.R. Beresford and F. Stajano, "Mix Zones: User Privacy in Location-Aware Services," *Proc. IEEE Ann. Conf. Pervasive Computing and Comm. Workshops (PERCOMW '04)*, Mar. 2004.
- [8] C. Bettini, X.S. Wang, and S. Jajodia, "Protecting Privacy against Location-Based Personal Identification," *Proc. Second VLDB Workshop Secure Data Management*, 2005.
- [9] "Rental Firm Uses GPS in Speeding Fine," *Chicago Tribune*, p. 9, July 2001.
- [10] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati, "K-Anonymity," *Secure Data Management in Decentralized Systems*, T. Yu and S. Jajodia, eds., Springer-Verlag, 2007.
- [11] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati, "Microdata Protection," *Secure Data Management in Decentralized Systems*, T. Yu and S. Jajodia, eds., Springer-Verlag, 2007.
- [12] E. Damiani, M. Anisetti, and V. Bellandi, "Toward Exploiting Location-Based and Video Information in Negotiated Access Control Policies," *Proc. Int'l Conf. Information Systems Security (ICISS '05)*, Dec. 2005.
- [13] T. D'Roza and G. Bilchev, "An Overview of Location-Based Services," *BT Technology J.*, vol. 21, no. 1, pp. 20-27, Jan. 2003.
- [14] M. Duckham and L. Kulik, "A Formal Model of Obfuscation and Negotiation for Location Privacy," *Proc. Int'l Conf. Pervasive Computing (PERVASIVE '05)*, May 2005.
- [15] M. Duckham and L. Kulik, "Dynamic & Mobile GIS: Investigating Change in Space and Time," *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 Trans. Mobile Computing*, vol. 7, no. 1, pp. 1-18, Jan. 2008.
- [17] Geographic Location/Privacy (Geopriv), <http://www.ietf.org/html.charters/geopriv-charter.html>, Sept. 2006.
- [18] G. Ghinita, P. Kalnis, and S. Skiadopoulos, "Privè: Anonymous Location-Based Queries in Distributed Mobile Systems," *Proc. Int'l World Wide Web Conf. (WWW '07)*, May 2007.
- [19] M. Gruteser and D. Grunwald, "Anonymous Usage of Location-Based Services through Spatial and Temporal Cloaking," *Proc. MobiSys '03*, 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*, vol. 22, no. 4, pp. 41-53, July 2005.
- [21] B. Ho and M. Gruteser, "Protecting Location Privacy through Path Confusion," *Proc. IEEE/CreateNet SecureComm '05*, Sept. 2005.
- [22] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Preserving Privacy in GPS Traces via Density-Aware Path Cloaking," *Proc. ACM Conf. Computer and Comm. Security (CCS '07)*, Oct. 2007.
- [23] M. Langheinrich, "Privacy by Design-Principles of Privacy-Aware Ubiquitous Systems," *Proc. Symp. Ubiquitous Computing (UBICOMP '01)*, Sept./Oct. 2001.
- [24] M. Langheinrich, "A Privacy Awareness System for Ubiquitous Computing Environments," *Proc. Symp. Ubiquitous Computing (UBICOMP '02)*, Sept./Oct. 2002.
- [25] J.-W. Lee, "Location-Tracing Sparks Privacy Concerns," *Korea Times*, <http://times.hankooki.com>, Nov. 2004.
- [26] Loopt, <http://www.loopt.com/>, Dec. 2008.
- [27] M.F. Mokbel, C.-Y. Chow, and W.G. Aref, "The New Casper: Query Processing for Location Services without Compromising Privacy," *Proc. Int'l Conf. Very Large Data Bases (VLDB '06)*, Sept. 2006.
- [28] P. Olofsson, *Probability, Statistics and Stochastic Processes*. John Wiley & Sons, Inc., 2005.
- [29] Privacy Rights Clearinghouse/UCAN, *A Chronology of Data Breaches*, <http://www.privacyrights.org/ar/ChronDataBreaches.htm>, 2006.
- [30] P. Samarati, "Protecting Respondents' Identities in Microdata Release," *IEEE Trans. Knowledge and Data Eng.*, vol. 13, no. 6, pp. 1010-1027, Nov./Dec. 2001.
- [31] H. Shin, V. Atluri, and J. Vaidya, "A Profile Anonymization Model for Privacy in a Personalized Location Based Service Environment," *Proc. Int'l Conf. Mobile Data Management (MDM '08)*, Apr. 2008.
- [32] G. Sun, J. Chen, W. Guo, and K.J.R. Liu, "Signal Processing Techniques in Network-Aided Positioning: A Survey of State-of-the-Art Positioning Designs," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 12-23, July 2005.
- [33] B. Thuraisingham, "Dependable Infrastructures and Data Managers for Sensor Networks," *Proc. IEEE Int'l Workshop Object-Oriented Real-Time Dependable Systems (WORDS '03)*, Oct. 2003.
- [34] B. Thuraisingham, "Directions for Security and Privacy for Semantic E-Business Applications," *Comm. ACM*, vol. 48, no. 12, pp. 71-73, Dec. 2005.
- [35] B. Thuraisingham, "Privacy Constraint Processing in a Privacy-Enhanced Database Management System," *Data and Knowledge Eng.*, vol. 55, no. 2, pp. 159-188, Nov. 2005.



[www.dti.unimi.it/ardagna](http://www.dti.unimi.it/ardagna).



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

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



can be found at <http://www.dti.unimi.it/decapita>.

**Sabrina De Capitani di Vimercati** is a professor in 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 International, California. She is a corecipient of the ACM-PODS'99 Best Newcomer Paper Award. More details about her research and background



**Pierangela Samarati** is a professor in 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, California, and a visiting researcher at Stanford University, California, and George Mason University, Virginia. 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 Special Interest Group on Security, Audit, and Control (SIGSAC). More details about her research and background can be found at <http://www.dti.unimi.it/samarati>.