Uniform Obfuscation for Location Privacy

Gianluca Dini and Pericle Perazzo[1]

University of Pisa,
Department of Information Engineering, via Diotisalvi 2,
56122 Pisa, Italy,
{g.dini, p.perazzo}@iet.unipi.it

Abstract. As location-based services emerge, many people feel exposed to high privacy threats. Privacy protection is a major challenge for such applications. A broadly used approach is *perturbation*, which adds an artificial noise to positions and returns an obfuscated measurement to the requester. Our main finding is that, unless the noise is chosen properly, these methods do not withstand attacks based on probabilistic analysis. In this paper, we define a strong adversary model that uses probability calculus to de-obfuscate the location measurements. Such a model has general applicability and can evaluate the resistance of a generic location-obfuscation technique. We then propose UniLO, an obfuscation operator which resists to such an adversary. We prove the resistance through formal analysis. We finally compare the resistance of UniLO with respect to other noise-based obfuscation operators.

Keywords: location-based services, privacy, obfuscation, perturbation, uniformity

1 Introduction

Recent years have seen the widespread diffusion of very precise localization technologies and techniques. The most known is GPS, but there are many other examples, like Wi-Fi fingerprinting, GSM trilateration, etc. The emergence of such technologies has brought to the development of *location-based services* (*LBS*) [3, 6, 10], which rely on the knowledge of location of people or things. The retrieval of people's location raises several privacy concerns, as it is personal, often sensitive, information. The indiscriminate disclosure of such data could have highly negative effects, from undesired location-based advertising to personal safety attempts.

A classic approach to the problem is to introduce strict access-control policies in the system [9, 16]. Only some trusted (human or software) entities will be authorized to access personal data. This access-control-based approach has a main drawback: if the entity does not need complete (or exact) information, it is a useless exposure of personal data. The "permit-or-deny" approach of access control is often too rigid. Some services require more flexible techniques which can be tailored to different user preferences.

Samarati and Sweeney [18] introduced the simple concept of *k-anonymity*: a system offers a k-anonymity to a subject if his identity is undistinguishable from (at least) $k-1$ other subjects. K-anonymity is usually reached by obfuscating data with some form of generalization. The methods based on k-anonymity [4, 10, 12] offer generally a high level of privacy, because they protect both the personal data and the subject's identity. However, they have some limitations:

- They do not permit the authentication of the subject and the customization of the service. Since they cannot identify the subject, some identity-based services like social applications or pay-services could not work.
- They are usually more complex and inefficient than methods based only on data obfuscation. This happens because their behavior must depend on a set of (at least) k subjects and not on a single subject only.
- They need a centralized and trusted obfuscator. In distributed architectures, such an entity may be either not present or not trusted by all the subjects.
- They are not applicable when the density of the subjects is too low. Obviously, if there are only 5 subjects in a system, they will never reach a 10-anonymity.

A simpler approach is *data-only obfuscation* [1, 15], whose aim is not to guarantee a given level of anonymity, but simply to protect the personal data. This is done by obfuscating data before disclosing it, in a way that it is still possible for the service provider to offer his service. Data obfuscation adds some artificial imperfection to information. The nature of such imperfection can fall into two categories [8]: *inaccuracy*, and *imprecision*. Inaccuracy concerns a lack of correspondence with reality, whereas imprecision concerns a lack of specificity in information. Deliberately introducing inaccuracy requires the obfuscation system to "lie" about the observed values. This can reduce significantly the number of assumptions the service can trust on. For this reason, the majority of obfuscation methods operates by adding imprecision, both by means of *generalization* or *perturbation* [5]. Generalization replaces the information with a value range which contains it, whereas perturbation adds random noise to it. We focus on the perturbation method. This method is both simple and efficient, and is often used to obfuscate data [14]. In spite of its simplicity, it requires to choose a suitable noise to effectively perturb data. In case of location data - and non-scalar data in general - such a problem is not trivial and should not be underrated. We found that if the noise is not chosen properly, perturbation will not resist to attacks based on statistical analysis. In particular, an obfuscation operator must offer a spacial *uniformity* of probability.

We present an analytical adversary model, which performs attacks based on statistical analysis. We show how such attacks can be neutralized by the property of uniformity. We present a metric for quantifying uniformity of an obfuscation system, called *uniformity index*. We further propose UniLO, an obfuscation operator for location data that introduces imprecision while maintaining accuracy. UniLO is simple and $\mathcal{O}(1)$-complex. It does not require a centralized and trusted obfuscator and can be seamlessly added to a distributed architecture as a building block. We show how UniLO offers a better uniformity with respect to other noise-based obfuscation operators. To the best of our knowledge, UniLO is the first obfuscation operator which offers guarantees on uniformity.

The rest of the paper is organized as follows. Section 2 introduces some basic concepts concerning the system model and the terminology. Section 3 formally describes the adversary model. Section 4 presents the UniLO operator in detail and its properties. Section 5 evaluates UniLO resistance by means of experimental results, and compares it to other obfuscation operators. Section 6 presents some examples of location-based services that can be built on UniLO operator. Section 7 explains some related works and analyzes differences and similarities with UniLO techniques. Finally, the paper is concluded in Section 8.

2 System Model

In our system, a *subject* is an entity whose location is measured by a *sensor*. A *service provider* is an entity that receives the subject's location in order to provide him with a *location-based service*. The subject applies an *obfuscation operator* to location information, prior to releasing it to the service provider. The obfuscation operator purposefully reduces the precision to guarantee a certain privacy level. Such a precision is defined by the subject and reflects his requirements in terms of privacy. The more privacy the subject requires, the less precision the obfuscation operator returns.

The subject is usually a person who has agreed to reveal - with some level of privacy - his location to one or more service providers. The service provider can be a human or a piece of software, depending on the kind of location-based service. For instance, a security service in an airport or in a train station often requires a human service provider. In contrast, in a customer-oriented service, for example, returning the nearest restaurant to the subject, the service provider may be a piece of software. The obfuscation operator can be applied to the data directly by the subject. Alternatively, a central obfuscator could be provided as well, serving several subjects at once.

For the sake of simplicity, the arguments and results we present in this paper refer to the two-dimensional case. However, they can be extended to the three-dimensional case in a straightforward way.

In the most general case, a *location measurement* is affected by an intrinsic error that limits its precision. Such error depends on several factors including the localization technology, the quality of the sensor, the environment conditions. Different technologies have different degrees of precision. For instance, the 68-th percentile of the error on a Garmin professional GPS receiver is 1.1 meters, on the iPhone's GPS is 8.6 meters, and on the iPhone's Wi-Fi localization system is 88 meters [21]. This implies that the location cannot be expressed as a geographical point but rather as a neighborhood of the actual location. We assume that locations are always represented as *planar circular areas* [1, 21], because it is a good approximation for many location techniques [17]. A location measurement (Fig. 1) can be defined as follows:

Definition 1 (Location measurement). *Let* $\mathbf{X}$ *be the* actual *position of the subject. A* location measurement *is a circular area* $A_m = \langle \mathbf{X}_m, r_m \rangle \subseteq \mathbb{R}^2$, *where* $\mathbf{X}_m$ *is the center of* A_m *and* r_m *is the radius, such that* $P\{\mathbf{X} \in A_m\} = 1$ (Accuracy Property).

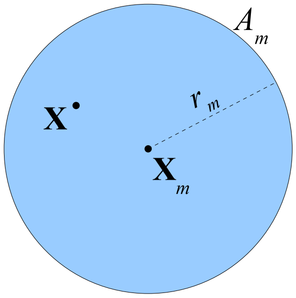


Fig. 1. Location measurement

The Accuracy Property guarantees that the location measurement actually contains the subject, or, equivalently, that the distance $\overline{\mathbf{X}\mathbf{X}_m}$ does not exceed r_m. The radius r_m specifies the precision of the localization technology, and we call it *precision radius*. Different technologies have different values for the precision radius. If a technology has a precision radius r_m, then a subject cannot be located with a precision better than r_m. We assume that r_m is constant over time. This means either that the precision does not change over time, or that r_m represents the worst-case precision.

A subject can specify his privacy preference in terms of *privacy radius* (r_p). If the subject specifies r_p, $r_p > r_m$, as his privacy radius, then he means that he wishes to be located with a precision not better than r_p. The task of an obfuscation operator is just to produce an obfuscated position $\mathbf{X}_p$, appearing to the provider as a measurement with precision r_p, worse than r_m. More formally, the obfuscation operator has to solve the following problem:

Problem 1 (Obfuscation). Let $\mathbf{X}$ be the actual position of a subject, $A_m = \langle \mathbf{X}_m, r_m \rangle$, be the location measurement and, finally, $r_p, r_p > r_m$, be his desired privacy radius. Transform, A_m into an *obfuscated measurement* (also called *privacy area*) $A_p = \langle \mathbf{X}_p, r_p \rangle$ such that the following properties hold:

1. (Accuracy) $P\{\mathbf{X} \in A_p\} = 1$
2. (Uniformity) *pdf* $(\mathbf{X}) : \mathbb{R}^2 \to \mathbb{R}$ (probability density function) as uniform as possible over A_p.

Property 1 guarantees that the obfuscated measurement actually contains the subject. Property 2 guarantees that the subject can be located everywhere in A_p with an almost-uniform probability. This property is particularly important because it prevents an adversary from determining areas that more likely contain the subject, and thus jeopardize the user privacy requirements. We will show how to quantify such a uniformity in Section 3.

With reference to Fig. 2, in order to produce an obfuscated measurement A_p, the obfuscation operator applies both an enlargement and a translation to the location measurement A_m. Intuitively, the operator enlarges the location measurement in order to decrease its precision and thus achieve the desired privacy level r_p. However, if A_m and A_p were concentric, determining the former from the latter would be trivial once the precision radius r_m is known. Therefore, the operator randomly selects a *shift vector* $\mathbf{d}$ and translates the enlarged measurement by $\mathbf{d}$, i.e., $\mathbf{X}_m + \mathbf{d} = \mathbf{X}_p$. Of course, the system has to keep the shift vector secret.

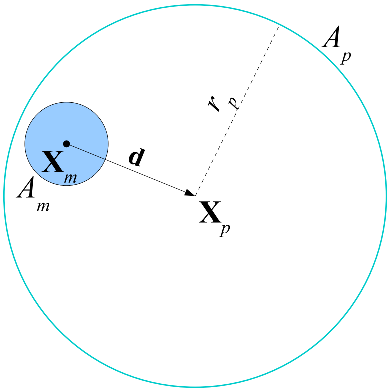


Fig. 2. Obfuscation and shift vector

The enlargement and translation operations must be such that, when composed, the resulting obfuscation satisfies the Accuracy and Uniformity Properties. Whereas the enlargement operation is straightforward, the translation operation is instead more subtle. As to the Accuracy Property, we state the following:

Proposition 1. *Given a location measurement A_m and an obfuscation $(r_p, \mathbf{d})$, the resulting obfuscated measurement A_p fulfills the Accuracy Property iff:*

$$\|\mathbf{d}\| \leq (r_p - r_m)$$

Proof. In order to guarantee the Accuracy Property, it is necessary and sufficient that $A_m \subset A_p$. Thus, with reference to Fig. 2, the distance between $\mathbf{X}_m$ and $\mathbf{X}_p$ must not exceed the difference between the precision radius and the privacy radius, i.e., $\|\mathbf{d}\| \leq (r_p - r_m)$.

3 Adversary Model and Uniformity Index

We assume the adversary knows the obfuscated measurement $\mathbf{X}_p$, the privacy radius r_p, and the precision radius r_m. She aims at discovering the actual subject's position $\mathbf{X}$. Since $\mathbf{X}$ cannot be known with infinite precision, the result of the attack will have a probabilistic nature.

Three kinds of information could help the adversary: (i) the probability density of the *measurement error*, which depends on the sensor's characteristics, (ii) the probability density of the *shift vector*, which depends on the obfuscation operator, and (iii) the probability density of the *population*, which depends on the map's characteristics. In the following, we will consider the population's density as *irrelevant* or, equivalently, *uniform*. This is a broadly used hypothesis in obfuscation systems [1]. In fact, landscape non-neutrality can be faced by means of complementary techniques, such as enlarging the privacy radius [2].

Basing on the measurement error's density and the shift vector's density, the adversary computes the *pdf* $f_{\mathbf{X}}(x,y)$ of the subject's position. After that, she defines a *confidence goal* $c \in (0,1]$ and computes the smallest area which contains the subject with a probability c:

Definition 2 (Smallest c-confidence area).

$$\hat{A}^c = \arg\min_{A \in \mathcal{A}^c} \{|A|\}$$

where:

$$\mathcal{A}^c = \left\{A | A \subseteq \mathbb{R}^2, P\{\mathbf{X} \in A\} = c\right\}$$
$$P\{\mathbf{X} \in A\} = \iint_A f_{\mathbf{X}}(x,y)\,\mathrm{d}x\mathrm{d}y$$

and $|A|$ indicates the size of A.

The adversary can find the smallest c-confidence area either analytically, by algebraic calculus, or statistically, by simulating many obfuscated measurements. $\hat{A}^c$ will cover the zones where $f_{\mathbf{X}}(x,y)$ is more concentrated. It is the result of the attack, and the adversary's most precise c-confidence estimation of the position. The smaller $\hat{A}^c$ is, the more precise is the adversary in locating the subject. A good obfuscation operator should keep $\hat{A}^c$ as larger as possible for every value of c. This is done by fulfilling the Uniformity Property. The best case occurs when the Perfect Uniformity Property is fulfilled, defined as follows:

Definition 3 (Perfect Uniformity Property). *An obfuscation operator fulfills the Perfect Uniformity Property iff $f_{\mathbf{X}}(x,y)$ is perfectly uniform over A_p.*

An obfuscation operator which fulfills such a property is *ideal.* It serves only for comparisons with real operators, and it is not realizable in the general case. This is because we cannot force a particular *pdf* inside A_p if we cannot control the *pdf* inside A_m, which depends on the measurement error.

Another way to state the Perfect Uniformity is the following:

Proposition 2. *A privacy area A_p fulfills the Perfect Uniformity Property iff:*

$$\forall A \subseteq A_p, P\left\{\mathbf{X} \in A\right\} = \frac{|A|}{|A_p|} \tag{1}$$

That is, each sub-area of A_p contains the subject with a probability proportional to its size. In such a case:

$$\left|\hat{A}^c\right| = c \cdot |A_p| \tag{2}$$

Otherwise:

$$\left|\hat{A}^c\right| < c \cdot |A_p| \tag{3}$$

The uniformity can be quantified by means of Eq. 3, by measuring how much, for a given c, $\left|\hat{A}^c\right|$ gets close to $c \cdot |A_p|$. We define the following *uniformity index* by fixing $c = 90\%$:

Definition 4 (Uniformity index).

$$\text{unif}\left(A_p\right) = \frac{\left|\hat{A}^{90\%}\right|}{90\% \cdot |A_p|}$$

The constant factor in the denominator is for normalization purposes. The uniformity index ranges from 0% (worst case), if the subject's position is perfectly predictable, to 100% (best case), if the subject's position is perfectly uniform. A uniformity index of 100% is necessary and sufficient for the Perfect Uniformity.

The uniformity index has a direct practical application. For example, if an obfuscation operator produces a privacy area of $400\,\text{m}^2$ with a uniformity index of 80%, the subject will be sure that an adversary cannot find his position (with 90% confidence) with more precision than $80\% \cdot 90\% \cdot 400 = 288\,\text{m}^2$. In other words, the uniformity index is proportional to the lack of precision of the attack.

4 UniLO

UniLO operator adds to $\mathbf{X}_m$ a shift vector $\mathbf{d} = (\mu\cos\phi, \mu\sin\phi)$ with the following probability densities (Fig. 3):

$$f\left(\phi\right) = \begin{cases} \frac{1}{2\pi} & \phi \in [0, 2\pi) \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$f\left(\mu\right) = \begin{cases} 2\mu/(r_p - r_m)^2 & \mu \in [0, r_p - r_m] \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

These equations aim at producing shift vectors with uniform spacial probability density, and magnitude less than or equal to $r_p - r_m$. This will greatly improve the uniformity of $f_{\mathbf{X}}\left(x, y\right)$. However, remind that $f_{\mathbf{X}}\left(x, y\right)$ depends even on the measurement error's density, over which we have no control. So it will not be perfectly uniform in the general case. UniLO fulfills the following properties:

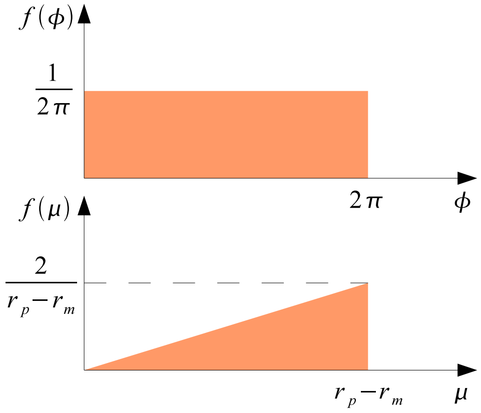


Fig. 3. ϕ and μ *pdf*s of a UniLO vector

Accuracy Property. The privacy area always contains the subject. We give a formal proof of this.

Uniformity Property. For $r_p/r_m \geq 10$, the uniformity index is above 81%. We will prove this by simulations, in Section 5.

Perfect Uniformity Property as $r_m \to 0$. With highly precise sensors, UniLO tends to be an ideal obfuscation operator. We give a formal proof of this.

Theorem 1. UniLO *fulfills Accuracy Property.*

Proof. By construction, $\|\mathbf{d}\| \leq r_p - r_m$. Hence, from Prop. 1, Accuracy holds.

Theorem 2. *As* $r_m \to 0$, UniLO *fulfills Perfect Uniformity Property.*

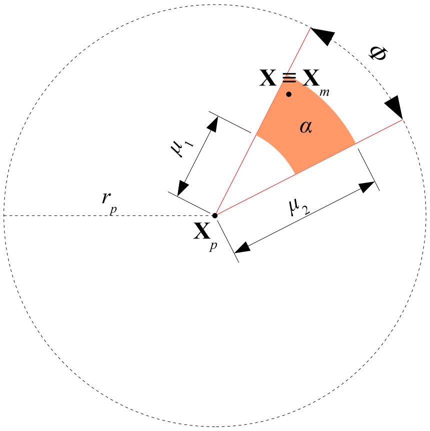


Fig. 4. Generic annular sector

Proof. If $r_m \to 0$, A_m will narrow to a point, with $\mathbf{X} \equiv \mathbf{X}_m$, and the probability density of the magnitude in Eq. 5 will become:

$$f(\mu) = \begin{cases} 2\mu/r_p^2 & \mu \in [0, r_p] \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Initially we prove that the hypothesis of Prop. 2 is satisfied for a generic annular sector α (Fig. 4). From Eqq. 4 and 6, and since $\mathbf{X} \equiv \mathbf{X}_m$:

$$\begin{aligned} P\{\mathbf{X} \in \alpha\} &= P\{\mathbf{X}_m \in \alpha\} \\ &= \int_0^{\Phi} \int_{\mu_1}^{\mu_2} \frac{2\mu}{r_p^2} \,\mathrm{d}\mu \,\frac{1}{2\pi} \mathrm{d}\phi \\ &= \frac{\Phi}{2} \frac{\left(\mu_2^2 - \mu_1^2\right)}{\pi r_p^2} \end{aligned}$$

Since the sizes of α and A_p are equal to:

$$\begin{aligned} |\alpha| &= \frac{\Phi}{2}\left(\mu_2^2 - \mu_1^2\right) \\ |A_p| &= \pi r_p^2 \end{aligned}$$

then:

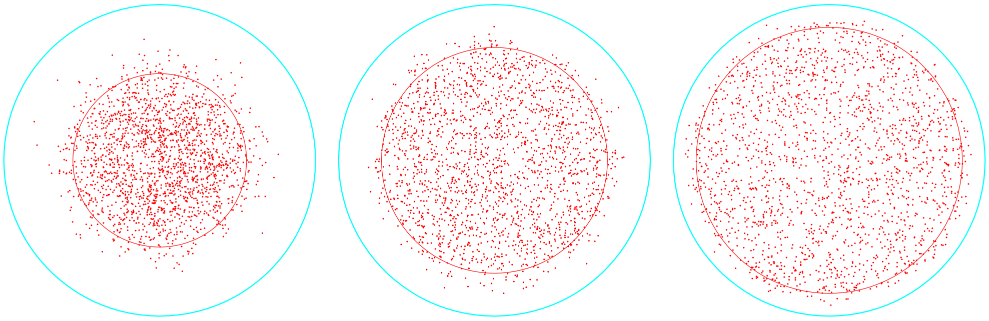
$$P\{\mathbf{X} \in \alpha\} = \frac{|\alpha|}{|A_p|}$$

If the hypothesis of Prop. 2 holds for a generic annular sector, it holds even for a composition of annular sectors, because the total size is the sum of the sizes, and the total probability is the sum of the probabilities. Since a generic $A \subseteq A_p$ can be partitioned in a set of infinitesimal annular sectors, the hypothesis of Prop. 2 holds for each $A \subseteq A_p$. Hence, Perfect Uniformity is satisfied.

It is worth remarking that UniLO operator protects a *single* obfuscated position. If the adversary can access many obfuscated positions at different times, as it happens in tracking systems, additional protection mechanisms must be deployed. In fact, if the subject does not move or moves slowly, the adversary could overlap the different privacy areas, thus reducing the uncertainty. A common countermeasure is to reuse the same shift vector every time [5]. If the subject does not move, the adversary will receive the same privacy area, and no overlap strategy will be possible.

5 Attack Resistance Analysis

UniLO has been implemented and used to obfuscate simulated location measurements. The error on the location measurements was assumed to follow a Rayleigh distribution, as it is usually done in GPS [13]. We truncated the distribution at $r_m = 3\sigma$, so that no sample falls outside A_m. Such truncated Rayleigh



(a) $r_p/r_m = 2$ (b) $r_p/r_m = 4$ (c) $r_p/r_m = 10$

Fig. 5. 2.000-sample simulations

distribution differs from the untruncated one for only 1.1% of samples. The tests aim at evaluating the uniformity of UniLO with respect to the ratio r_p/r_m (*radius ratio*).

Figure 5 shows the statistical distribution of $\mathbf{X}$ in A_p of 2.000 UniLO samples for different values of the radius ratio. They give a first visual impression about the uniformity of UniLO. We note that the distribution tends to be perfectly uniform as $r_p/r_m \to \infty$. The inner areas are $\hat{A}^{90\%}$.

We compared UniLO with other common obfuscation noises:

- A Rayleigh noise (i.e. gaussian X - gaussian Y), used for modeling 2-dimensional measurement errors. The Rayleigh distribution is truncated at $r_p - r_m$, in order to fulfill Accuracy Property. The σ parameter is fixed at $(r_p - r_m)/3$.
- A gaussian-μ noise (i.e. uniform angle - gaussian magnitude), used by Krumm to perturb GPS data [14]. The gaussian distribution is truncated at $r_p - r_m$, in order to fulfill Accuracy Property. The σ parameter is fixed at $(r_p - r_m)/3$.
- A uniform-μ noise (i.e. uniform angle - uniform magnitude). This is the simplest two-dimensional noise.

Figure 6 shows the uniformity indexes of the noises. Each uniformity index estimation was obtained by means of 50 million samples. As we told in Section 4, UniLO offers a uniformity index above 81% for $r_p/r_m \geq 10$. We can see how UniLO performs better than all the other noises for all the radius ratii. In particular, gaussian-magnitude and Rayleigh-magnitude noises are particularly bad for obfuscating. We believe this is the reason why Krumm needed a surprisingly high quantity of noise ($\sigma = 5\,\mathrm{Km}$) to effectively withstand inference attacks [14].

6 Service Examples

UniLO operator has the advantage to be transparent to the service provider, in the sense that a privacy area has the same properties as an ordinary measurement area. A software service provider designed for receiving non-obfuscated

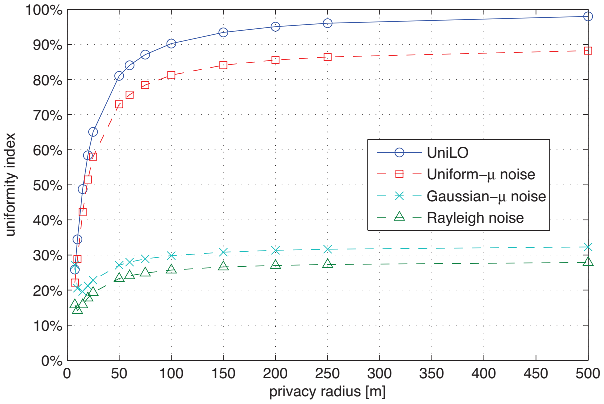


Fig. 6. unif (A_p) with $r_m = 5\,\mathrm{m}$

inputs can be seamlessly adapted for receiving UniLO-obfuscated inputs. The following subsections describe some examples of services which can be deployed over UniLO operator.

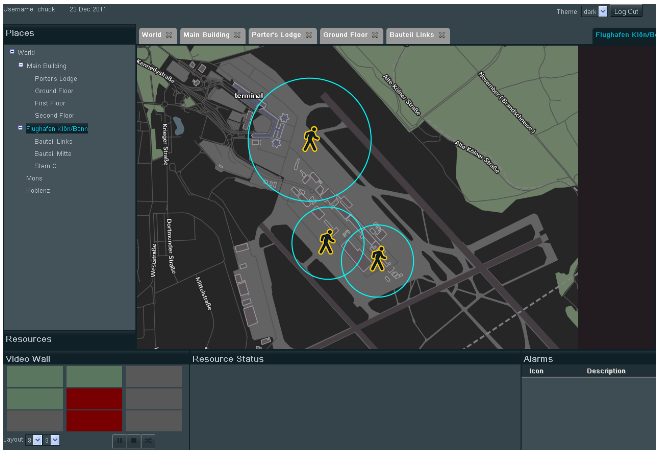
6.1 Employee localizer

The aim is to retrieve the instantaneous locations of a set of employees to better coordinate work operations. Before giving their consensus, employees specify their privacy radii. A software service provider displays the locations on the monitor of a human operator, in the form of circles on a map. Each circle is larger or smaller depending on the privacy radius. The privacy radius may depend on context-based rules. For example, an employee may require a high privacy radius when standing in some zones of the map and a small one when standing in others. Figure 7 shows a screenshot of such a service, taken from a practical implementation.

6.2 Find the near friends

This is a social application, in which the users share their obfuscated positions with their friends. Alice wants to find out which of her friends are in her proximity. We define "being in the proximity of Alice" as "being at a distance of 400 meters or less from Alice". In this case Alice is the service provider and her friends are the subjects. While Alice knows its own position, the locations of her friends are obfuscated. Suppose Bob is one of Alice's friends. Since Alice does not know his exact location, the question "is Bob in my proximity?" will necessarily have a probabilistic answer, like "60% yes, 40% no".

The problem can be modeled as depicted in Fig. 8. Alice builds a circle centered on its position and with 400 meters of radius (*proximity circle*, A), and



Username: chuck 23 Dec 2011

Theme: dark

Fig. 7. Employee localizer screenshot

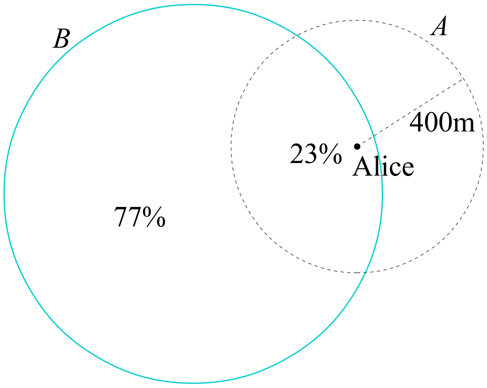


Fig. 8. Find the near friends

computes the intersection between that circle and the privacy circle of Bob (B). If Bob is inside this intersection, he will be in Alice's proximity. The probability that such an event happens is:

$$P\,\{\text{Bob is in Alice's proximity}\} = \iint_{A\cap B} f\,(x,y)\;\mathrm{d}x\mathrm{d}y \tag{7}$$

Alice can numerically compute such an integral to find out the probability. If the privacy area of Bob can be assumed as perfectly uniform, the Eq. 7 will become:

$$P\,\{\text{Bob is in Alice's proximity}\} = \frac{|A\cap B|}{|B|}$$

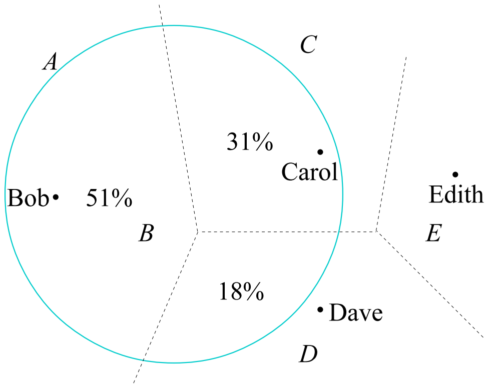


Fig. 9. Find the nearest taxi

In the figure, such probability is 23%. The service provider performs this calculus only for each friend whose $\mathbf{X}_p$ is nearer than $r_p + 400\,\mathrm{m}$. The others have no intersection, and thus 0% probability. Alice finds an answer like the following:

- Bob is in the proximity with 23% probability.
- Carol with 10% probability.
- Dave with 100% probability.
- All the others with 0% probability.

6.3 Find the nearest taxi

Alice calls a taxi and releases her obfuscated GPS position in order to speed-up the procedure. The taxi company knows the positions of the available taxis. Then, it finds the one which is probabilistically the nearest to Alice, and forwards the request to it. In this way, only the taxi driver needs to know Alice's exact position.

The problem can be modeled as depicted in Fig. 9, by means of a Voronoi diagram. Each region of the diagram corresponds to a taxi. Let us call the taxi drivers Bob (region B), Carol (C), Dave (D) and Edith (E). If Alice is inside B, Bob's will be the nearest taxi, and so on. Fortune's algorithm [11] can compute the Voronoi diagram in $\mathcal{O}(n \log n)$ time, where n is the number of taxis. The taxi company obtains the probabilities by simply integrating $f(x, y)$ over the intersections between the privacy area and the Voronoi regions. If the privacy area of Alice can be assumed as perfectly uniform, the integral becomes a simple area ratio, like in Subsection 6.2. In the figure, the taxi company will obtain the following probabilities:

- Bob's taxi is the nearest with 51% probability.
- Carol's taxi with 31% probability.
- Dave's taxi with 18% probability.
- All the others with 0% probability.

The taxi company will then forward the request to Bob.

7 Related Works

Conway and Strip published a seminal work about general-purpose database-oriented obfuscation methods [5]. The authors introduced two obfuscation approaches that, with some generalization, have been used until today: *value distortion*, which perturbs value with a random noise; and *value-class membership*, which partitions the whole value domain in classes, and discloses only the class where the value is in.

Gruteser and Grunwald first approached k-anonymity problem in location-based services. The proposed solution involves the subdivision of the map in static quadrants with different granularities [12]. Mascetti et al. proposed an obfuscation method that divides the map in quadrants like [12], but it does not aim at k-anonymity [15]. It focuses only on data obfuscation and proximity services.

Duckham and Kulik took a radically different approach, that models a map as an adjacency graph, where the vertices represent zones of the map and the edges the adjacency between two zones [7]. A graph modelization is more powerful in some applications, because it can model obstacles, unreachable zones and hardly viable passages through edge costs. The obfuscation method reveals a set of nodes where the subject could be. Proximity services are realized by means of Dijkstra-like algorithms. Shokri et al. took a similar approach, and involves also the anonymization of the subjects [20]. A drawback is that a graph-based description of the map must be available, and shared between the subjects and the service providers. Calculating a graph model of a geographic map that is both simple and accurate may be not trivial. Another drawback is that the proximity services are not based on simple and efficient Voronoi diagrams (cfr. Section 6), but they have to involve more complex Dijkstra-like algorithms.

Ardagna et al. proposed a set of obfuscation operators that perturb the location: radius enlargement, radius restriction, center shift [1]. These operators transform a measurement area into an obfuscated one. To the best of our knowledge, this is the most similar work to our approach, but it contains relevant differences with respect to UniLO in the initial requirements and the final results:

- The subject's actual location could be outside the obfuscated area. This happens in case of radius reduction or center shift operators. Thus, the obfuscation introduces inaccuracy which does not allow the service provider to offer some services, like those described in Section 6. In contrast, UniLO always guarantees that the obfuscated area contains the subject.
- The quantity of privacy is measured by a parameter, called *relevance*, which is quite unintuitive. Final users prefer parameters they can easily understand such as the *privacy radius* used by UniLO. If a user specifies a privacy radius of 100 m, then he means that he wishes to be located with a precision not better than 100 m. Relevance has not a 1-to-1 relationship with the privacy radius: the same relevance corresponds to a small privacy radius if the location technology is precise, or to a larger one if is imprecise.

- The resistance against attacks relies on the fact that the system chooses the obfuscation operators at random. However, the adversary can make probabilistic hypothesis on them. This possibility is not investigated. De facto, the adversary is assumed to be unaware of the obfuscation method. This is an optimistic assumption, which features a form of security by obscurity that should be avoided [19].

8 Conclusions and Future Works

We have proposed UniLO, an obfuscation operator for location data, which adds a special random noise which maximizes probability uniformity. UniLO is simple and $\mathcal{O}(1)$-complex. We have presented an adversary model which performs statistical-based attacks. We have shown that the property of uniformity neutralizes such attacks. We have proved the resistance of UniLO in terms of uniformity, through both formal analysis and experimental results. To the best of our knowledge, UniLO is the first obfuscation operator which offers guarantees on uniformity.

The work leaves space for extensions to noncircular or nonplanar location measurements, extensions for tracking systems, and extensions to offer multiple contemporaneous levels of privacy.

Acknowledgment

This work has been supported by the EU-funded Integrated Project PLANET “PLAtform for the deployment and operation of heterogeneous NETworked cooperating objects,” and the Network of Excellence CONET “Cooperating Objects Network of Excellence.”

References

1. Ardagna, C.A., Cremonini, M., De Capitani di Vimercati, S., Samarati, P.: An obfuscation-based approach for protecting location privacy. IEEE Transactions on Dependable and Secure Computing 8(1), 13–27 (Jan 2011)
2. Ardagna, C.A., Cremonini, M., Gianini, G.: Landscape-aware location-privacy protection in location-based services. Journal of Systems Architecture 55(4), 243–254 (Apr 2009)
3. Barkuus, L., Dey, A.: Location-based services for mobile telephony: a study of users privacy concerns. In: Proceedings of the INTERACT 2003, 9th IFIP TC13 International Conference on Human-Computer Interaction. pp. 709–712 (Jul 2003)
4. Beresford, A.R., Stajano, F.: Location privacy in pervasive computing. IEEE Pervasive Computing 2(1), 46–55 (Jan 2003)
5. Conway, R., Strip, D.: Selective Partial Access to a Database. In: Proceedings of the 1976 Annual Conference. pp. 85–89. ACM (1976)
6. D’Roza, T., Bilchev, G.: An overview of location-based services. BT Technology Journal 21(1), 20–27 (Jan 2003)

7. Duckham, M., Kulik, L.: A formal model of obfuscation and negotiation for location privacy. In: Gellersen, H.W., Want, R., Schmidt, A. (eds.) Proceedings of the PERVASIVE 2005: 3rd International Conference on Pervasive Computing. vol. 3468, pp. 152–170. Springer Berlin / Heidelberg (2005)
8. Duckham, M., Mason, K., Stell, J., Worboys, M.: A formal approach to imperfection in geographic information. Computer, Environment and Urban Systems 25, 89–103 (1999)
9. Duri, S., Gruteser, M., Liu, X., Moskowitz, P., Perez, R., Singh, M., Tang, J.M.: Framework for security and privacy in automotive telematics. In: Proceedings of the 2nd International Workshop on Mobile Commerce. pp. 25–32. ACM (2002)
10. Espinoza, F., Persson, P., Sandin, A., Nyström, H., Cacciatore, E., Bylund, M.: GeoNotes: Social and navigational aspects of location-based information systems. Tech. Rep. T2001/08, Swedish Institute of Computer Science (SICS) (May 2001)
11. Fortune, S.: A sweepline algorithm for voronoi diagrams. In: Proceedings of the Second Annual ACM SIGACT/SIGGRAPH Symposium on Computational Geometry. pp. 313–322. SCG '86, ACM (1986)
12. Gruteser, M., Grunwald, D.: Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. Proceedings of the MobiSys 2003: 1st International Conference on Mobile Systems, Applications and Services pp. 31–42 (2003)
13. Hofmann-Wellenhof, B., Lichtenegger, H., Collins, J.: Global Positioning System: Theory and Practice. Springer (2001)
14. Krumm, J.: A survey of computational location privacy. Personal and Ubiquitous Computing 13(6), 391–399 (2008)
15. Mascetti, S., Bettini, C., Freni, D., Wang, X.S., Jajodia, S.: Privacy-Aware Proximity Based Services. In: Proceedings of the MDM 2009: 10th International Conference on Mobile Data Management: Systems, Services and Middleware. pp. 31–40. IEEE (2009)
16. Myles, G., Friday, A., Davies, N.: Preserving privacy in environments with location-based applications. IEEE Pervasive Computing 2(1), 56–64 (Jan 2003)
17. Pal, A.: Localization algorithms in wireless sensor networks: Current approaches and future challenges. Network Protocols and Algorithms 2(1), 45–74 (2010)
18. Samarati, P., Sweeney, L.: Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Tech. rep., Computer Science Laboratory SRI International (1998)
19. Schneier, B.: Secrecy, security, and obscurity (May 2002), www.schneier.com/crypto-gram-0205.html
20. Shokri, R., Freudiger, J., Jadliwala, M., Hubaux, J.P.: A distortion-based metric for location privacy. In: Proceedings of the 8th ACM Workshop on Privacy in the Electronic Society. pp. 21–30. WPES '09, ACM (2009)
21. Zandbergen, P.A.: Accuracy of iPhone locations: A comparison of assisted GPS, WiFi and cellular positioning. Transactions in GIS 13(s1), 5–26 (Jun 2009)