Protecting Users' Anonymity in Pervasive Computing Environments

Linda Pareschi Daniele Riboni Claudio Bettini
D.I.Co., University of Milano
via Comelico, 39
20135 Milano, Italy
{pareschi,riboni,bettini}@dico.unimi.it

Abstract

The large scale adoption of adaptive services in pervasive and mobile computing is likely to be conditioned to the availability of reliable privacy-preserving technologies. Unfortunately, the research in this field can still be considered in its infancy. This paper considers a specific pervasive computing scenario, and shows that the application of state-of-the-art techniques for the anonymization of service requests is insufficient to protect the privacy of users. A specific class of attacks, called shadow attacks*, is formally defined and a set of defense techniques is proposed. These techniques are validated through the use of a simulator and an extensive set of experiments.*

1 Introduction

The proliferation of cheap sensing technologies, powerful portable devices, and wireless networks has recently enabled the diffusion of new classes of context-aware pervasive services; i.e., services that adapt themselves to the current situation of users in a pervasive computing environment.

Server-side adaptation involves the communication to the service provider of private information about users, such as their current activity and location, personal data, interests, preferences, and, in some cases, physiological data. Hence, one of the most challenging issues in this research area is to devise effective techniques for preserving users' privacy while guaranteeing a satisfactory quality of service as a result of the adaptation process. Indeed, it has been shown that simply hiding users' explicit identifiers (e.g., SSN) may not be sufficient to guarantee privacy, because in several cases the real user identity can be inferred from the other data communicated to the service provider.

To this aim, various approaches have been proposed for privacy preservation in pervasive and mobile computing ([7, 21, 17, 8]), mainly based on *access control* (e.g., [22]) or *anonymization* (e.g., [10, 3]). While access control provides a very robust solution for controlling the disclosure of private information, we argue that a solution based solely on access control is not well suited to several kinds of services. Consider, for instance, a location-based service (LBS). In this case, if location is the data to protect and the service provider is considered an untrusted entity, completely negating access to the user's location data would determine the impossibility of providing the service at all. A more flexible solution to this problem is provided by anonymization techniques. In this case, the data to be protected is communicated to the service provider after having been partially generalized or suppressed, and consequently the identity of the actual issuer becomes indistinguishable in a set of k potential issuers. This kinds of techniques are generally referred to as k-anonymity techniques [10, 16, 5].

As argued in [5], the soundness of privacy protection techniques strongly depends on the assumptions about the knowledge available to possible adversaries. Techniques for enforcing k-anonymity in LBS have focused on protection against attacks performed on the basis of data included in users' service *requests*. In this paper, we present a novel class of attacks on k-anonymity, that can be performed on the basis of *service responses* and *users' behavior* (sets of users' context data acquired by the adversary) as a result of the received responses. In order to exemplify a similar attack, we introduce a pervasive computing scenario that will be used throughout this paper.

Consider the pervasive system of a gym (called *PerGym* and sketched in Figure 1) in which users wear a smart watch that collects context data from body-worn sensors to continuously monitor data such as user's position (acquired through a user-side indoor positioning system), the used equipments (through RFID), and physiological parameters. These data are communicated from users to the *virtual trainer* service of the gym included in a request to obtain suggestions for the next exercise. Since physiological data are particularly sensitive (because they can reveal important details about a person's health status), they need to

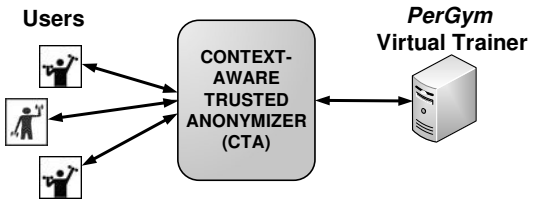


Figure 1. The *PerGym* scenario

be anonymized. For this purpose, requests are sent through an encrypted channel to a context-aware trusted anonymizer (CTA) in charge of enforcing k-anonymity. Context data are kept up-to-date on the CTA by periodical updates through an encrypted channel. Suppose that the gym system is considered untrusted by its users (hence, from a privacy perspective, the gym system is considered a possible adversary). The gym system is also able to collect a subset of the context data known by the CTA; in particular, it can continuously monitor the users' positions through a server-side positioning system. Since it knows users' identities, their position, and the gym map, it is anytime aware of who is using a given equipment.

We claim that in a similar scenario state-of-the-art k-anonymity techniques are insufficient to preserve users' privacy. For instance, suppose that user u_1 submits a request r_1 for the next exercise to be performed. Data included in r_1 are partially generalized by the CTA (which also replaces u_1's identity with a *pseudo-id* p_1) in order to make u_1 indistinguishable between k potential issuers. Then, the resulting request r_1' is sent to the gym service provider, which responds with a list of possible equipments, and corresponding exercises, communicated by service response s_1. If u_1 under pseudo-id p_1 starts to use a suggested equipment that was suggested only to p_1, since the *PerGym* system can monitor part of the behavior of the potential issuers (i.e., their movements and activity in the pervasive space) and knows their identities, it can associate with high probability the pseudo-id p_1 to the actual issuer u_1 of r_1' . Hence, k-anonymity is broken as a consequence of this attack, that we call *shadow attack*.

To the best of our knowledge, shadow attacks have never been addressed before. They have some similarities with attacks considering multiple requests issued by the same user, for which the notion of k-anonymity has been extended to *historical k-anonymity* [6]. These attacks mainly rely on identifying a set of requests as issued by the same user, and then on using the information sent with these requests. For example, considering the location and time of the user as reported in the requests, it is possible to reduce the number of potential issuers to those that actually were observed in all those locations at those times. In shadow attacks, even if pseudo-ids are continuously changed in order to avoid the identification of a trace of requests from the same issuer, privacy can still be at risk, since the attack is not based on analyzing the next request of the user, but on correlating the observation of actual users with the content of service responses. Clearly, shadow attacks depend on the knowledge of service responses, which are not taken into account by attacks on historical k-anonymity.

Shadow attacks can be applied only to those scenarios in which service responses can influence the future behavior of users. However, we believe that this is the case for many pervasive computing scenarios, in which users issue service requests for having access to physical resources (e.g., in [13, 19]), receiving directions (e.g., in [2, 9]) and suggestions (e.g., proximity advertising [1, 12]).

The main contributions of this paper can be summarized as follows:

- We show that state-of-the-art techniques for privacy protection through anonymity are insufficient in pervasive computing scenarios;
- We formalize a new kind of attack, called shadow attack;
- We propose defense techniques for shadow attacks and present an experimental evaluation.

The paper is structured as follows: in Section 2 we provide preliminary information about privacy protection; in Section 3 we exemplify and formalize shadow attacks; in Section 4 we propose defense techniques; in Section 5 we report experimental results in a simulated scenario; Section 6 concludes the paper.

2 Preliminaries

A privacy threat is generally intended as the possibility that an adversary associates the user's identity to *private information (PI)*. This association is denoted *sensitive association (SA)*. In order to prevent the release of the *SA* it is possible to modify the released data in order to increase the uncertainty about the user's identity or about the private information. The uncertainty on the user identity is called *anonymity*: it has been introduced for data base systems ([20]) and then adapted to LBS services ([4]). The main idea of anonymity is to make the actual issuer of a LBS request indistinguishable in a set, called *anonymity set*, of potential issuers. The cardinality of the *anonymity set* determines the degree k of anonymity achieved for a given request.

In order to violate the user's privacy, i.e. discovering the SA, an adversary can access external knowledge (e.g., positioning systems, telephone books) that joined with data included in a request can restrict the set of candidate issuers.

Categories of data that joined with external knowledge can increase the probability of reconstructing the *SA* are defined *quasi-identifier (QI)*([18]). Clearly, which elements of a request act as QI strongly depend on the external knowledge available to the adversary. In most research papers on privacy, spatio-temporal information are considered QI since the adversary often happens to know the position of users.

Most k-anonymity techniques are based on the generalization/suppression of QI data, and on the replacement of the user's unique identifier with a *null* value or with a *pseudo-id*. Hence, each request r is transformed by a third part (in our scenario the CTA) into a request r' with the identity and the QIs components appropriately modified to enforce k-anonymity.

3 Attacking anonymity with *shadow attacks*

In this section we show that state-of-the-art anonymity techniques for privacy protection are insufficient when applied to a pervasive computing scenario, and we formalize shadow attacks.

3.1 Attacking anonymity in the *PerGym* scenario

Consider the following example:

Example 1 *Suppose that the CTA enforces k-anonymity with $k = 10$, and that, in a given time granule, three users u_1, u_2 and u_3 submit a service request r_1, r_2, and r_3, respectively. Hence, before forwarding the three requests to the service provider, the CTA anonymizes the requests by generalizing the quasi-identifiers in the request (i.e., time, user's location and activity, age) to make the issuers indistinguishable in a set of at least 10 users. The other service parameters (e.g., physiological data) are not generalized. However, the latter data are assumed not to be quasi-identifiers, then they cannot help the adversary in associating the requests with their actual issuers. Suppose that, as a result of generalization, u_1, u_2 and u_3 belong to the same anonymity set $\{u_1, u_2, \ldots, u_{10}\}$. Upon receiving (or intercepting) the three generalized requests r'_1, r'_2, and r'_3, the adversary cannot definitely associate a request r'_i with its issuer u_j; in fact, from the adversary's perspective, each request r'_i has $1/10$ probability of having been issued by user u_j (for simplicity, we assume the adversary is performing a* uniform attack *[5] on k-anonymity).*

The above example shows how k-anonymity can be profitably used for preserving users' privacy from attacks performed by considering service request parameters and context data (e.g., users' location and activity). However, a different class of attacks on k-anonymity needs to be considered; i.e., attacks based on the analysis of *service responses* and *users' behavior* as a consequence of the received service responses. We call this class of attacks *shadow attacks*, because for performing them the adversary must (electronically or, in extreme cases, physically) *shadow* the behavior of the possible issuers belonging to the anonymity set. The following example illustrates a shadow attack.

Example 2 *Suppose that the service provider, based on location of the user in the gym, gender, age, and physiological data, suggests – by means of the service responses – the following gym equipments (e.g., exercise bikes, rowing machines) to the three users (obviously, the adversary does not know the association between user identities and service responses; in fact, in the request the real user identity is substituted by a* pseudo-id*):*

- *equipments e_1, e_2 and e_3 are suggested to user u_1;*
- *equipments e_1, e_4 and e_5 are suggested to user u_2;*
- *equipments e_5, e_6 and e_7 are suggested to user u_3.*

We recall that the suggested gym equipments are sets *of equipments (e.g., "one of the exercise bikes"); the system does not suggest an individual equipment (e.g., "the first exercise bike in room R1"). For the sake of simplicity, also suppose that an authorization mechanism is adopted for ensuring that each user can only use an equipment that the service provider suggested to her in a given time window (this restrictive assumption will be relaxed in the definition of defense techniques). Then, as a consequence of the received service responses, suppose that the three users decide to use the following equipments:*

- *an equipment e_1 is used by user u_1;*
- *an equipment e_1 is used by user u_2;*
- *an equipment e_6 is used by user u_3.*

Since the system can continuously monitor the behavior of users in the gym, the adversary can perform a shadow attack to decrease the provided anonymity level, or even to definitely associate a user with her service request – and, hence, with the private information included in the request. In particular:

- *Since users u_1 and u_2 decided to use the same equipment e_1 the adversary cannot be sure if u_1 issued r'_2 and u_2 issued r'_1 or vice versa; i.e., they are still indistinguishable among themselves. However, the provided anonymity level for u_1 and u_2 decreases from 10-anonymity to 2-anonymity.*
- *Since user u_3 decided to use an equipment e_6 that was not suggested to any other user in the anonymity set, the adversary can unambiguously associate u_3 with*

her request r'_3; hence, even enforcing k-anonymity at the time of the service request, the private information included in the request performed by u_3 is disclosed as a consequence of a shadow attack.

The above example is illustrative of the peculiar characteristics of shadow attacks, which must be taken into account for devising effective defense techniques:

- Privacy protection depends on the behavior of the user with respect to service responses. In fact, if user u_1 would have chosen to use an equipment e_3 it would have been unambiguously associated with her service request, since e_3 was not suggested to any other user in her anonymity set;
- The behavior of users can impact on the privacy level of other users. In fact, if user u_2 would have chosen to use an equipment e_4, then user u_1 would have been unambiguously associated with her service request (by elimination);
- Service responses can influence the future behavior of users; hence, a malicious service provider (or an intermediate entity that can maliciously modify service responses) can produce responses that facilitate the association between a user and her service request (e.g., suggesting the use of resources that were not suggested to any other user in the anonymity set).

Given their peculiar characteristics, shadow attacks can be applied to a wide range of pervasive computing scenarios. In fact, several pervasive computing services are aimed at suggesting the use of physical resources (e.g., meeting rooms, printers, projectors) to users on the basis of context.

3.2 Formal definition of *shadow attacks*

As pointed out in [5], in order to devise sound defense techniques and evaluate their properties, a framework for formally defining concepts like *attack*, *external knowledge*, and *defense function* is needed. For this reason, in this paper we adopt the formal framework proposed by Bettini et Al. in [5], and extend it to define shadow attacks and defense techniques against them.

We recall that, given a set of requests and generalized requests R, a set of users' identities I, and the external knowledge Γ available to an adversary, an attack is defined in [5] as follows:

Definition 1 (from [5]) *An* attack *based on knowledge Γ is a function $Att_\Gamma : R \times I \rightarrow [0,1]$ such that for each generalized request r',*

$$\sum_{i \in I} Att_\Gamma(r', i) = 1.$$

Hence, the value of $Att_\Gamma(r', i)$ represents the probability of the individual i to be the issuer of the generalized request r', as inferred from an attack performed on the basis of Γ.

It must be noted that Definition 1 models attacks performed on the basis of generalized service requests and external knowledge; service responses and users' behavior as a consequence of the received responses are not explicitly taken into account. Even if these latter data can be considered part of the external knowledge available to the adversary, in order to ease the definition and evaluation of shadow attacks and their defense strategies we extend the framework proposed in [5] with the following definitions:

Definition 2 *Given a service s, a generalized request r' to that service, the external knowledge Γ available to the adversary, and the corresponding anonymity set $\Lambda = Anon_\Gamma(r')$, we call*

$$\Phi_{s,\Lambda,\tau}$$

the set of responses to requests for s by users in Λ during time interval τ.

Definition 3 *Given the external knowledge Γ available to the adversary, the anonymity set Λ, and the time interval τ', we call* shadowed behavior $\Psi_{\Gamma,\Lambda,\tau'}$ *the set of context data about users in Λ during τ' acquired by the adversary.*

Finally, we can formally define a shadow attack as follows:

Definition 4 *A* shadow attack *based on service responses $\Phi_{s,\Lambda,\tau}$ and shadowed behavior $\Psi_{\Gamma,\Lambda,\tau'}$ (τ' immediately follows τ) is a function $SAtt_{\Phi_{s,\Lambda,\tau},\Psi_{\Gamma,\Lambda,\tau'}} : R \times I \rightarrow [0,1]$ such that for each generalized request r',*

$$\sum_{i \in I} SAtt_{\Phi_{s,\Lambda,\tau},\Psi_{\Gamma,\Lambda,\tau'}}(r', i) = 1.$$

Similarly to Att_Γ, the function $SAtt_{\Phi_{s,\Lambda,\tau},\Psi_{\Gamma,\Lambda,\tau'}}$ models the probability of an individual i to be the issuer of the generalized request r'. Actually, shadow attacks can be considered a particular class of attacks in which external knowledge includes service responses and shadowed behavior of users:

Property 1 *Given Definition 1, a shadow attack is an attack in which*

$$\Gamma \supseteq \{\Phi_{s,\Lambda,\tau}, \Psi_{\Gamma,\Lambda,\tau'}\}.$$

In the following, we illustrate how the example presented in Section 3.1 can be formally described:

Example 3 *Continuing Examples 1 and 2, we have that:*

- *I corresponds to the identities of users in the gym;*

- s *is the* virtual trainer *service suggesting the next exercise on the basis of context;*
- *the private information is physiological data;*
- $\Lambda = \{u_1, u_2, \ldots, u_{10}\}$*;*
- τ *is the time interval during which* u_1*,* u_2 *and* u_3 *submit their service requests;*
- Γ *includes the exact location of users in the gym, the gym map, and users' identities;*
- τ' *is the time interval during which* u_1*,* u_2 *and* u_3 *move to a new equipment in order to perform a new exercise, after having received suggestions from* s*;*
- $\Phi_{s,\Lambda,\tau}$ *corresponds to the first itemized list in Example 2;*
- $\Psi_{\Gamma,\Lambda,\tau'}$ *corresponds to the second itemized list in Example 2;*
- *the values of* $SAtt_{\Phi_{s,\Lambda,\tau},\Psi_{\Gamma,\Lambda,\tau'}}$ *with respect to users* u_1*,* u_2 *and* u_3 *and generalized requests* r'_1*,* r'_2 *and* r'_3 *are reported in Table 1.*

$SAtt_{\Phi_{s,\Lambda,\tau},\Psi_{\Gamma,\Lambda,\tau'}}(r'_i, u_j)$	u_1	u_2	u_3
r'_1	0.5	0.5	0
r'_2	0.5	0.5	0
r'_3	0	0	1

Table 1. Probability of an individual u_j to be the issuer of generalized request r'_i as inferred from a shadow attack (from Example 3)

4 Defense techniques

In this section we propose defense techniques for shadow attacks.

Estimating privacy threats In order to apply effective protection techniques for shadow attacks, it is necessary to estimate the level of privacy threat deriving from possible users' behaviors. To this aim, we call $p(E_{u_i,\omega_j})$ the probability of the event *"user* u_i *chooses alternative* ω_j*"*, where $\omega_j \in \Omega_{a_i}$, and Ω_{a_i} is the set of alternatives proposed by service response a_i. The success of a shadow attack also depends on the conditional probability $p(\Psi'_{u_i,\omega_j} \subseteq \Psi_{\Gamma,\Lambda,\tau'}|E_{u_i,\omega_j})$ that the adversary gets to know which of the alternatives has been chosen by users (we call Ψ'_{u_i,ω_j} the set of context data sufficient to characterize the behavior consequent to E_{u_i,ω_j}).

For simplicity, in the following we assume that any of the alternatives $\omega_j \in \Omega_{a_i}$ proposed by a service response a_i has the same probability of being chosen by user u_i; i.e., $p(E_{u_i,\omega_j}) = \frac{1}{|\Omega_{a_i}|}$. Moreover, we also assume that the adversary can precisely know which alternative has been chosen by users; i.e., $p(\Psi'_{u_i,\omega_j} \subseteq \Psi_{\Gamma,\Lambda,\tau'}|E_{u_i,\omega_j}) = 1$. Those assumptions are not realistic in most scenarios; however, our privacy protection techniques can be applied to any other probability distribution.

Finally, we call p_{u_i,Ψ_j} the probability that an adversary correctly associates the identity u_i to her service request by observing the behavior Ψ_j consequent to the choice of alternative ω_j (e.g., observing that u_i is using equipment e_j).

Example 4 *Consider the anonymity set and the service responses illustrated in Example 2. Suppose that the three service responses are cached by the CTA, and sent at the same time to the three users. Table 2 shows the values of* p_{u_i,Ψ_j} *for users* u_i *and behavior* Ψ_j *(that corresponds to the use of equipments* e_j*). For instance, the probability* p_{u_1,Ψ_1} *that an adversary correctly associates the request of user* u_1 *to her identity by a shadow attack is* 0.83 *if she would choose equipment* e_1 *(since it was also suggested to* u_2*), while it is* 1 *if she would choose equipments* e_2 *or* e_3 *(that were suggested to her alone).*

p_{u_i,Ψ_j}	Ψ_1	Ψ_2	Ψ_3	Ψ_4	Ψ_5	Ψ_6	Ψ_7
u_1	0.83	1	1	×	×	×	×
u_2	0.83	×	×	1	0.83	×	×
u_3	×	×	×	×	0.83	1	1

Table 2. Privacy threats table PTT of Example 4

Even if usability issues are out of the scope of this paper, we point out that users are informed about privacy threats corresponding to possible behaviors – by means of user-friendly interfaces – in order to support them in choosing privacy-conscious behaviors. Moreover, our privacy protection strategy includes the suppression by the CTA of those service responses that are associated with a high privacy threat.

Definition 5 *We call the table providing values of* p_{u_i,Ψ_j} *for users belonging to* Λ *and their possible behaviors as a result of service responses* privacy threats table PTT.

Enlarging τ Obviously, the more users in Λ submit service requests, the more difficult is for an adversary to associate service responses to users' behavior. Hence, a possible defense technique consists in choosing a time interval

τ (during which users' requests are cached by the CTA before being anonymized and sent to the service provider) that is sufficiently large to ensure that – in most cases – a relevant portion of users in Λ submit a service request. However, since the value of τ impacts on the service response time, this technique must be applied with caution, in order to avoid the introduction of an excessive delay in service provision.

Generating fake requests Similarly to what proposed in [11] for privacy protection in location-based services, a further defense strategy for shadow attacks consists in the use of *fake requests*, i.e., fictitious requests – seemingly submitted by users in Λ – sent by the CTA to artificially decrease the values of p_{u_i,Ψ_j} in PTT.

In general, the use of fake requests should be avoided, since – from the service provider perspective – it may imply a high overhead for the infrastructure. Hence, in our proposal the CTA generates fake requests only if an adversary has high probability of associating real identities with service requests (i.e., if most values in PTT are close to 1). Unfortunately, an analysis of PTT before sending fake requests is unsafe. In fact, sending fake requests after real ones would allow the adversary to easily recognize fakes. As a consequence, the algorithm we devised generates fake requests only if the ratio between real requests (i.e., requests submitted by users in Λ) and k is below a certain threshold f.

Example 5 *Consider the scenario depicted in Example 2, and suppose that, for that scenario, the chosen value of f is 0.5. Hence, the CTA generates 2 fake requests r'_4 and r'_5 that are sent – together with the real ones and in random order – to the service provider. The service provider responds as follows:*

- *equipments e_1, e_5 and e_6 are suggested as a response to r'_4;*
- *equipments e_1, e_2 and e_5 are suggested as a response to r'_5.*

The service provider responds to the real requests as stated in Example 2. The corresponding PTT is shown in Table 3.

p_{u_i,Ψ_j}	Ψ_1	Ψ_2	Ψ_3	Ψ_4	Ψ_5	Ψ_6	Ψ_7
u_1	0.31	0.5	1	×	×	×	×
u_2	0.31	×	×	1	0.31	×	×
u_3	×	×	×	×	0.31	0.5	1

Table 3. Privacy threats table PTT with the generation of fake requests (Example 5)

As can be seen from Table 3, the use of fake requests determines a relevant improvement in the values of PTT; in fact, the best (lower) value of p_{u_i,Ψ_j} decreases from 0.83 to 0.31 for each user.

Continuously updating privacy threats estimates As it is shown by the following example, the values and shape of the privacy threats table PTT changes when the behavior of users in Λ, as observed by an adversary on the basis of Γ, can be associated to one or more service responses.

Example 6 *Continuing Example 5, suppose that user u_1 decides to use equipment e_2. Then, PTT is changed, as shown in Table 4. As a consequence, behavior Ψ_5 becomes safer than Ψ_1 for user u_2 (they were equally safe before observation of Ψ_1 of user u_1).*

p_{u_i,Ψ_j}	Ψ_1	Ψ_2	Ψ_3	Ψ_4	Ψ_5	Ψ_6	Ψ_7
u_2	0.33	×	×	1	0.31	×	×
u_3	×	×	×	×	0.31	0.5	1

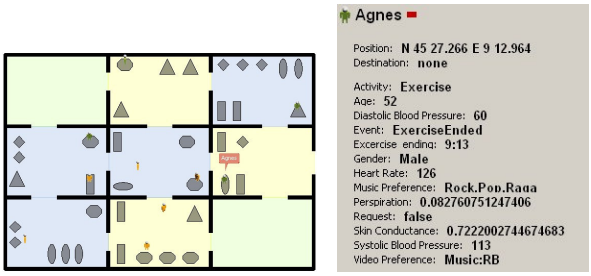
Table 4. Updated PTT (Example 6)

In order to exploit this feature, our defense strategy includes a mechanism based on asynchronous notifications for informing users about changes in privacy threats as a consequence of other users' behavior.

5 Experimental evaluation

In this section we present an experimental evaluation of our proposal. In order to evaluate the effectiveness of our defense techniques against shadow attacks, we have simulated the *PerGym* scenario using the *Siafu* [14] context simulator. Since in this set of experiments we concentrated on protection against shadow attacks, and not against generic attacks on k-anonymity, we simplified our assumptions, and we assumed that the only *quasi-identifier* included in service requests is user location. Hence, the generalization function takes into account only location. The development of an algorithm for multidimensional k-anonymity in pervasive services will be the subject of future work.

Simulation of the *PerGym* scenario We have simulated the *PerGym* scenario by means of the Siafu context simulator [14]. Siafu provides facilities for describing the physical characteristics of an environment (e.g., walls, furnishings, devices, equipments), as well as for modeling the behavior (e.g., movements, activities) of users within that environment as sets of context data that change with time. Figure 2 shows a screenshot of the *PerGym* environment and context data generated by Siafu.



(a) *PerGym* environment (b) Context data

Figure 2. ***PerGym*** **model**

Experiments have been performed with a population of 300 users moving in a gym of $3600\,m^2$. The gym comprises 20 types of equipments. At generation time, each user is associated to a set of context data randomly initialized. Context data include personal information (e.g. name, gender, age), current location (in the form of latitude - longitude degrees), current activity (e.g., exercising, resting), video and music preferences (chosen among values organized in hierarchies), and physiological data (e.g., blood pressure, heart rate, skin conductance).

Context data influence the future behavior of users; exercises last on average 14 minutes, and are followed by on average 1 minute of rest. When a user finishes an exercise, she has 90% probability of sending a request asking for the next exercise; in the other case she leaves the gym. When a user leaves the gym, the arrival of a new user is simulated.

k-anonymity and generalization Our first experiment was aimed at evaluating the degree of generalization of users' location on the basis of the level k of k-anonymity. For performing generalization we have adopted the *DicotomicArea* algorithm [15], both for its good performance and ease of implementation.

Figure 3 shows that the average area of the generalized users' locations increases almost linearly with k. The "staircase" behavior of the plot is due to the behavior of the *DicotomicArea* algorithm, as explained in [15].

In our scenario, the average generalized area is approximately $30\,m^2$ with $k = 10$; with values of k greater than 20, the average generalized area is greater than $90\,m^2$. With values of k greater than 40, the average generalized area becomes too large for the kind of services envisioned in the *PerGym* scenario.

Defense based on τ In a second set of experiments we evaluated the effectiveness of the defense technique based on an enlargement of the time granule τ during which users' requests are cached by the CTA before being generalized and sent to the service provider. The greater τ, the more

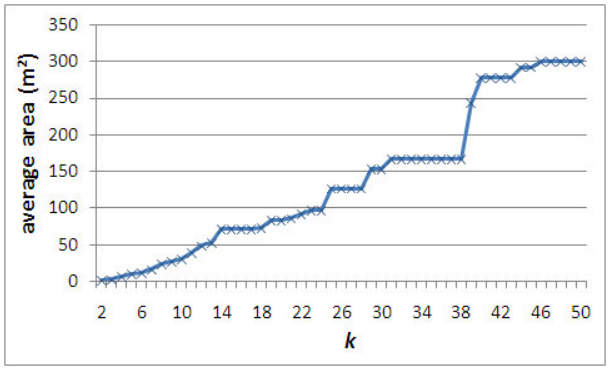


Figure 3. Degree of generalization of location on the basis of k

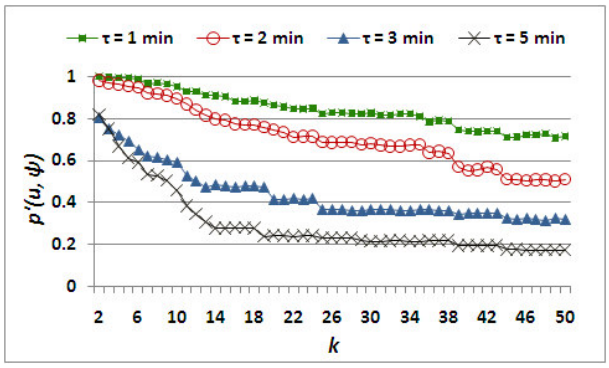


Figure 4. Average probability p'_{u_i,Ψ_j} depending on τ and k

difficult is for the adversary to associate requests to their actual issuers. However, since τ introduces delays in service responses, the choice of τ is constrained by the service characteristics. In our case, since we assume that users can take a few minutes of rest between an exercise and the subsequent, we have performed our experiments with values of τ between 1 minute and 5 minutes. We recall that each user submits a service request on average every 15 minutes.

Figure 4 shows the average probability p'_{u_i,Ψ_j} that an adversary unambiguously associates the identity of user u_i to her actual generalized service request r'_i by observing the behavior Ψ_j (that corresponds to the use of a suggested equipment j); i.e., $SAtt(r'_i, u_i) = 1$. As expected, results show that the larger τ, the lower the privacy threat. In particular, given a $k = 10$ degree of k-anonymity, values of p'_{u_i,Ψ_j} are greater than 0.89 with $\tau \leq 2$ minutes; they are less than 0.59 with $\tau \geq 3$ minutes.

We also studied the probability p''_{u_i,Ψ_j} that an adversary

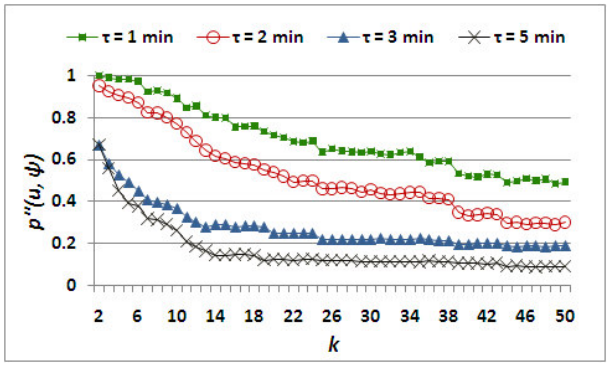


Figure 5. Average probability p''_{u_i,Ψ_j} depending on τ and k

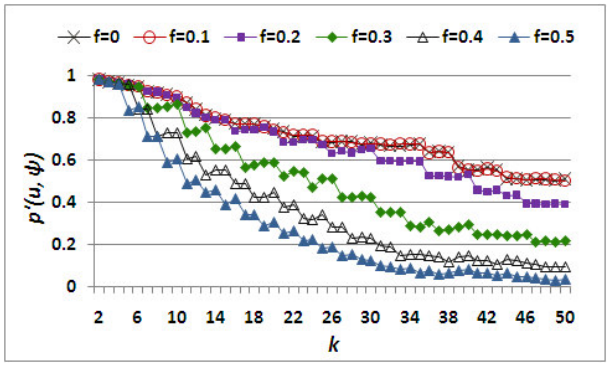


Figure 6. Average probability p'_{u_i,Ψ_j} depending on f and k

unambiguously associates the identity of user u_i to her actual generalized service request r'_i in the case u_i chooses the *safest* alternative j. The safest alternative is the one (or one of the ones) having the lower value of p_{u_i,Ψ_j} in the privacy threat table PTT. Figure 5 shows that p''_{u_i,Ψ_j} is also strongly influenced by τ. Low levels of privacy threat ($p''_{u_i,\Psi_j} < 0.2$) can be obtained only with $\tau \geq 5$ minutes and $k > 10$. The choice of this value of τ is not well suited to our scenario, since it would introduce an excessive delay in service responses. Hence, the next set of experiments aimed at evaluating the use of fake requests for providing better privacy protection while using a more reasonable value of τ.

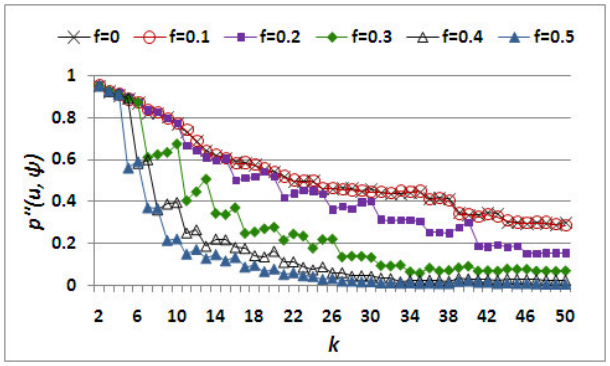


Figure 7. Average probability p''_{u_i,Ψ_j} depending on f and k

Defense based on fake requests When the choice of τ is constrained by service requirements, privacy protection can be enhanced by the use of fake requests. In this set of experiments we have chosen a value of $\tau = 2$ minutes, that corresponds to an average additional delay of 1 minute for each service request. We have evaluated both p'_{u_i,Ψ_j} and p''_{u_i,Ψ_j} with different values of f (between 0 and 0.5) and k. We recall that our privacy protection technique includes the generation of fake requests if the ratio between real requests (submitted by users in Λ) and k is below a given threshold f.

Figures 6 and 7 report experimental results, and show that the use of fake requests decreases the level of privacy threat. In particular, with the use of fake requests and $f = 0.5$ the average value of p'_{u_i,Ψ_j} with $k = 10$ is 0.6 (it is 0.89 without the use of fake requests). If we consider p''_{u_i,Ψ_j}, results show that low levels of privacy threat can be obtained with $k = 10$ and $\tau = 2$ minutes if our defense technique based on fake requests is applied with $f = 0.5$.

6 Conclusions and future work

In this paper we have shown that state-of-the-art k-anonymity techniques for privacy protection are insufficient when applied to many pervasive computing scenarios. We have formalized shadow attacks and proposed defense techniques, which have been experimentally evaluated in a simulated environment.

Some of the most relevant issues we are currently investigating are *a)* the definition of a comprehensive measure of privacy, *b)* an extension of our privacy protection techniques to support multidimensional k-anonymity for context-aware services, and *c)* an extension to support the *dynamic case*, i.e., when an adversary is able to reconstruct the sensitive association by means of requests issued by the same user in different time intervals.

References

[1] L. Aalto, N. Göthlin, J. Korhonen, and T. Ojala. Bluetooth and WAP Push Based Location-Aware Mobile Advertising System. In *Proceedings of the Second International Conference on Mobile Systems, Applications, and Services (MobiSys 2004)*. USENIX, 2004.

[2] M. Arikawa, S. Konomi, and K. Ohnishi. Navitime: Supporting Pedestrian Navigation in the Real World. *IEEE Pervasive Computing*, 6(3):21–29, July–Sept 2007.

[3] A. R. Beresford and F. Stajano. Location Privacy in Pervasive Computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.

[4] C. Bettini, S. Mascetti, and X. S. Wang. Privacy Protection through Anonymity in Location-based Services. *Handbook of Database Security: Applications and Trends*, pages 509–530, 2007.

[5] C. Bettini, S. Mascetti, X. S. Wang, and S. Jajodia. Anonymity in Location-based Services: Towards a General Framework. In *Proceedings of the 8th International Conference on Mobile Data Management (MDM)*, pages 69–76. IEEE Computer Society, 2007.

[6] C. Bettini, X. S. Wang, and S. Jajodia. Protecting Privacy Against Location-Based Personal Identification. In *Proceedings of the 2nd Workshop on Secure Data Management (SDM)*, volume 3674 of *LNCS*, pages 185–199. Springer, 2005.

[7] R. Campbell, J. Al-Muhtadi, P. Naldurg, G. Sampemane, and M. Mickunas. Towards Security and Privacy for Pervasive Computing. In *Proceedings of International Symposium on Software Security*, volume 2609 of *Lecture Notes in Computer Science*, pages 1–15. Springer, 2003.

[8] H. S. Cheng, D. Zhang, and J. G. Tan. Protection of Privacy in Pervasive Computing Environments. In *ITCC '05: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II*, pages 242–247. IEEE Computer Society, 2005.

[9] K. Cheverst, N. Davies, K. Mitchell, and A. Friday. Experiences of Developing and Deploying a Context-aware Tourist Guide: the GUIDE Project. In *Proceedings of MOBICOM'00*, pages 20–31. ACM, 2000.

[10] M. Gruteser and D. Grunwald. Anonymous Usage of Location-based Services through Spatial and Temporal Cloaking. In *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services (MobiSys)*, pages 31–42. USENIX Association, 2003.

[11] H. Kido, Y. Yanagisawa, and T. Satoh. Protection of Location Privacy using Dummies for Location-based Services. In *ICDEW '05: Proceedings of the 21st International Conference on Data Engineering Workshops*. IEEE Computer Society, 2005.

[12] S. Kurkovsky and K. Harihar. Using Ubiquitous Computing in Interactive Mobile Marketing. *Personal and Ubiquitous Computing*, 10(4):227–240, 2006.

[13] C. Lee and A. Helal. Context Attributes: An Approach to Enable Context-awareness for Service Discovery. In *Proceedings of the 2003 Symposium on Applications and the Internet (SAINT 2003)*, pages 22–30. IEEE Computer Society, 2003.

[14] M. Martin and P. Nurmi. A Generic Large Scale Simulator for Ubiquitous Computing. In *Third Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services, 2006 (MobiQuitous 2006)*, pages 1–3. IEEE Computer Society, 2006.

[15] S. Mascetti and C. Bettini. A Comparison of Spatial Generalization Algorithms for LBS Privacy Preservation. In *Proceedings of the 1st International Workshop on Privacy-Aware Location-based Mobile Services (PALMS)*. IEEE Computer Society, 2007.

[16] M. F. Mokbel, C.-Y. Chow, and W. G. Aref. The New Casper: Query Processing for Location Services without Compromising Privacy. In *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB)*, pages 763–774. VLDB Endowment, 2006.

[17] G. Myles, A. Friday, and N. Davies. Preserving Privacy in Environments with Location-based Applications. *IEEE Pervasive Computing*, 2(1):56–64, 2003.

[18] P. Samarati. Protecting Respondents' Identities in Microdata Release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):1010–1027, 2001.

[19] M. Samulowitz, F. Michahelles, and C. Linnhoff-Popien. CAPEUS: An Architecture for Context-Aware Selection and Execution of Services. In *Proceedings of the Third International Working Conference on Distributed Applications and Interoperable Systems (DAIS)*, volume 198 of *IFIP Conference Proceedings*, pages 23–40. Kluwer, 2001.

[20] L. Sweeney. k-Anonymity: a Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (IJUFKS)*, 10(5):557–570, 2002.

[21] R. Wishart, K. Henricksen, and J. Indulska. Context Obfuscation for Privacy via Ontological Descriptions. In *Proceeding of First International Workshop on Location- and Context-Awareness (LoCA)*, volume 3479 of *Lecture Notes in Computer Science*, pages 276–288. Springer, 2005.

[22] M. Youssef, V. Atluri, and N. R. Adam. Preserving Mobile Customer Privacy: an Access Control System for Moving Objects and Customer Profiles. In *MDM '05: Proceedings of the 6th International Conference on Mobile Data Management*, pages 67–76. ACM Press, 2005.