Towards Privacy Protection in a Middleware for Context-awareness

(Short Paper)

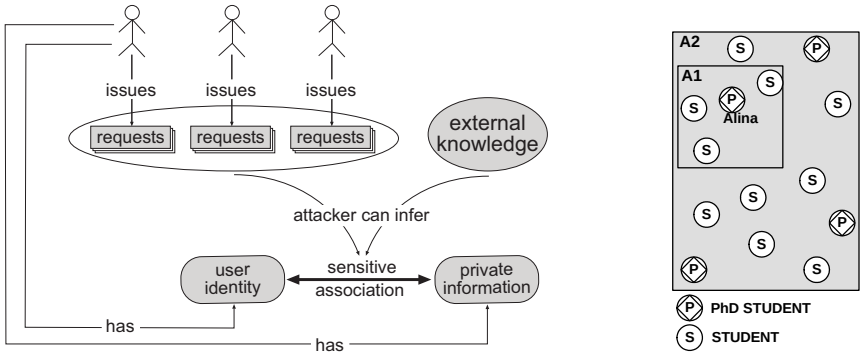
Linda Pareschi, Daniele Riboni, Sergio Mascetti, and Claudio Bettini

D.I.Co., University of Milan
via Comelico, 39, I-20135, Milan, Italy
{pareschi,riboni,mascetti,bettini}@dico.unimi.it

Abstract. Privacy is recognized as a fundamental issue for the provision of context-aware services. In this paper we present work in progress regarding the definition of a comprehensive framework for supporting context-aware services while protecting users' privacy. Our proposal is based on a combination of mechanisms for enforcing context-aware privacy policies and k-anonymity. Moreover, our proposed technique involves the use of stereotypes for generalizing precise identity information to the aim of protecting users' privacy.

1 Introduction

The recent proliferation of powerful mobile devices, wireless networks, and sensing technologies has enabled new classes of context-aware services, e.g., services that adapt themselves to the current situation of the user. However, since adaptation involves the communication to the service provider of user's private information such as her location, activity, and profile data, in order to be accepted by final users these services must be supported by mechanisms for preserving privacy. In the literature about privacy in location-based services (LBS) it has been shown that simply hiding users' explicit identifiers (e.g., user's name) may not be sufficient to guarantee privacy, since the user's identity can be possibly inferred from the other information sent to the service provider. To this aim, several approaches have been proposed for enforcing access control, or for guaranteeing anonymity. On the basis of the experience we have acquired in the last years while working on a framework for context-awareness [1], and on privacy in LBS [2], we argue that a satisfactory comprehensive solution for privacy in context-awareness is still missing. As a matter of fact, we believe that a solution based solely on access control is unsuitable for many services, since simply negating the access to a given context data (e.g., location) would determine the impossibility of providing the service at all (e.g., a LBS). On the other hand, work on anonymity for context-aware services has generally concentrated on solely location, while the set of data that can be useful for adaptation is much wider. In this paper we present work in progress regarding the definition of a comprehensive framework for privacy protection in context-awareness.



(a) Privacy threat (b) Scenario

Fig. 1. Privacy threat and use-case scenario

External knowledge assumptions The privacy issue we are addressing, shown in Figure 1(a), consists in preventing the attacker from inferring the *sensitive association*, i.e. the association between the user identity and the *private information* (PI) of that user. According to the knowledge an attacker can acquire from external sources, data included in requests may increase his ability to reconstruct the sensitive association (these data are called *quasi-identifiers* (QI) [3]).

Techniques for privacy preservation strongly depend on which knowledge an hypothetical attacker can access. Therefore, in order to provide effective mechanisms for protecting the user privacy, external knowledge assumptions must be formalized with respect to the application logic of the required service and the privacy threat to be contrasted.

2 Requirements and proposed solution

In order to illustrate how we intend to address privacy issues, we consider the following running example:

Example 1. Consider a context-aware service providing points of interest ($POIs$) to mobile users, depending on data such as location, profile, and interests. Suppose that a hypothetical user Alina – a PhD student – is submitting a request to that service using her GPS-enabled smart phone, in order to find a bookshop.

The service presented in the above example is representative of a large number of mobile services, in which adaptation is performed considering not only location, but a wide set of context data.

Since we claim that formal knowledge assumptions are fundamental for defining a privacy preservation technique, we couple the running example described above with the following illustrative knowledge assumption:

- Γ_C: the attacker can acquire knowledge about context data (including spatio-temporal information and profile data), either directly from external sources, or inferring them from other public data.

Assuming Γ_C, each context data included in a requests can possibly act as QI.

According to this external knowledge assumption, in order to preserve Alina's privacy it is necessary to hide either her precise location, and her profile and interests. Approaches based on *access control techniques* may determine the negation of that kind of data to the service provider. However, such a solution would not be suitable to the above mentioned service – and, in general, to any context-aware service– since spatio-temporal and context data are essential for selecting a set of resources that can be potentially interesting to the user.

A better solution would consist in the application of *obfuscation techniques*; i.e., in providing generalized data instead of their precise values. For instance (see Figure 1(b)), instead of communicating the exact information regarding Alina, it would be possible to provide the POI service with an area $A1$ including her precise location (e.g., the block where she currently is), her status (i.e., a PhD student), and her interests I. This solution would allow the service provider to select a set of POIs that are actually close to Alina's interests and location. However, this approach does not prevent the attacker from identifying the issuer of the request. In fact, Alina could be identified if the attacker has the knowledge of Alina being the only PhD student in the area $A1$ having interests I.

Therefore, we have identified the following requirements for a privacy preservation middleware addressed to users of context-aware services:

(a) A flexible mechanism for disclosing obfuscated context data, still guaranteeing a given level of privacy, is mandatory;

(b) Personal data acting as QI must be generalized in a meaningful way in order to allow the service provider to adapt the service;

(c) When location is not the only context data to be provided, the obfuscation process must determine the level of generalization of each context data in order to fulfill either privacy and adaptation requirements.

In order to protect the sensitive association between the user's identity and her private information, obfuscation techniques can be applied either to QIs, to PIs, or to both QIs and PIs. However, as a first effort towards the definition of a comprehensive framework for privacy protection in context-awareness, in this work we focus our investigation on the first approach.

2.1 *k*-Anonymity for users of context-aware services

In order to address requirement *(a)*, we adopt the *k-anonymity* technique already introduced in database systems ([4]) and applied to LBS [5, 2]. The intuition behind *k-anonymity* consists in making the user indistinguishable in a set of k potential issuers. According to Γ_C, any context data may act as QI; therefore, in order to enforce *k-anonymity*, any context data must be considered in the generalization process.

Generalization depends on the semantics and representation of context data; for instance, if the data is represented by one (or more) numerical values (e.g., GPS coordinates), the exact data is generalized to an interval (or to an area). On the contrary, if the data is represented by a specific value v belonging to

a taxonomy T, it can be generalized to an ancestor of v in T. For example, given a possible stereotype hierarchy, the exact stereotype of Alina (i.e., *PhD student*) can be generalized to its parent *student*. Since the problem of optimal multidimensional anonymization is NP-hard, we plan to adopt an approximation algorithm for addressing multidimensional *k-anonymity*.

2.2 Identity generalization through stereotypes

In order to protect the user's privacy, any data that uniquely identifies the user must be removed from the user request. Hence, the most commonly adopted solution consists in substituting the *user-id* (together with any possible user personal data) with an anonymous *pseudo-id*, which can be used by the application logic for performing tasks such as authentication and session management. However, since in such a solution any reference to the user personal data are lost, the service provider is no longer able to customize the service according to data such as the user age, gender, and formal education, which are considered very relevant for accurately personalizing services.

According to the requirement *(b)*, we propose to extend the *pseudo-id* approach by the use of *stereotypes* [6] for generalizing the user's identity. Stereotypes are useful abstractions for characterizing users' demographic data, personalities, and goals. Our solution allows the application logic not only to perform authentication and session management, but also to customize the service according to relevant (generalized) personal data, while preserving anonymity.

2.3 Context-aware privacy policies

The generalization of context data obviously impacts on the quality of adaptation. Indeed, even if *refinement* techniques can be used for improving the service response, it could happen that, in order to achieve the desired anonymity level, context data become too general to provide the service at an acceptable quality level [7]. In order to address this issue, we allow users to declare privacy policies about the generalization of single context data, depending on the context itself:

Example 2. Alina declared the following privacy policies:

p1: If *activity==working* Then *anonymity-level:=high*

p2: If *activity==shopping* Then *anonymity-level:=low*

p3: If *activity==walking* Then *provide-accurate-location*

Alina asks the POI server for a bookshop. Since she is currently strolling for shopping, policies **p2** and **p3** hold: then, her desired anonymity level is set to a low value (corresponding to 4-anonymity), and the anonymizer provides a quite accurate location (in order to select resources that are actually close to her). Hence, in order to enforce Alina's policies, the anonymizer generalizes Alina's exact stereotype to *student* for obtaining a smaller area $A1$ containing 4 potential issuers (see Figure 1(b)), and communicates those data to the POI server.

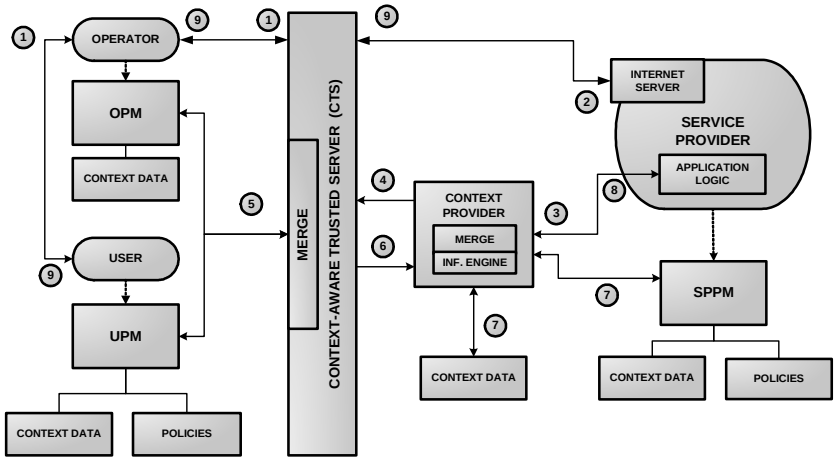


Fig. 2. Architecture overview

Moreover, for each context data the user can define the maximum granularity of generalization on the basis of context (e.g., *if I am walking, then generalize location to at most 500m*). Since it is not always possible to conciliate these constraints with the desired level of anonymity, a conflict resolution mechanism is adopted for choosing if either guaranteeing a lower anonymity level, generalizing the context data to a higher granularity, or not providing the service at all.

2.4 Architecture overview

In order to illustrate how privacy preserving techniques can be integrated into an architecture for context-awareness, in this section we consider an extension of the CARE middleware [1]. CARE supports the acquisition of context data from different sources, the reasoning with this data based on distributed policies, and the reconciliation of possibly conflicting information. Figure 2 depicts an extension of CARE to support privacy: the module devoted to apply our privacy preservation techniques (called CONTEXT-AWARE TRUSTED SERVER (CTS)) acts as an intermediary between the user trusted domain (left-hand side) and the rest of the world (right-hand side). The user trusted domain is constituted by the USER PROFILE MANAGER (UPM) – which manages user policies, and context data explicitly provided by the user or acquired from user-side sensors – and by the NETWORK OPERATOR PROFILE MANAGER (OPM), which manages context data such as users' location and network status.

Numbers in Figure 2 represent the data flow upon a user request: each user request (1) is filtered by the CTS, which transforms the user ID into a *pseudoID* – which is used to identify the user request and to perform authentication – and removes any QI, before forwarding the request (2). Next (3), the service provider asks for the context data it needs for adapting the service to a central module called CONTEXT PROVIDER, which forwards the request to the CTS (4).

The CTS retrieves user's privacy policies, and distributed context data from the user trusted domain, it merges context data solving possible conflicts, and generalizes them according to the privacy policies (5). Then (6), it sends those data – together with the (possibly generalized) user stereotype – to the CONTEXT PROVIDER, which merges them with context data retrieved from the SERVICE PROVIDER PROFILE MANAGER (SPPM) and external context sources (7), and evaluates service provider policies, thus obtaining the aggregated context data that are communicated to the application logic (8). Finally (9), the application logic adapts the service and communicates the service response to the CTS, which forwards it to the user. Note that, in order to avoid eavesdropping, any communication between the CTS, the service provider, and the CONTEXT PROVIDER is encrypted with public key cryptography.

3 Future work

This preliminary investigation has highlighted several important requirements that will be the object of our future research efforts. In the following we mention the most important aspects we are planning to consider: *i)* in order to allow users to define their own privacy preferences depending on context, a sufficiently expressive language for context-aware privacy preferences is needed; *ii)* in order to estimate the quality of service according to the chosen privacy policies, a mechanism for measuring the trade-off between adaptation quality and anonymization degree must be devised; *iii)* an efficient mechanism for multidimensional context data generalization must be defined; *iv)* effective privacy techniques must be applied to a *dynamic case*, i.e., when an attacker is able to reconstruct the sensitive association by means of requests issued by the same user in different time intervals.

References

1. Agostini, A., Bettini, C., Cesa-Bianchi, N., Maggiorini, D., Riboni, D., Ruberl, M., Sala, C., Vitali, D.: Towards Highly Adaptive Services for Mobile Computing. In: Proc. of IFIP TC8 Conf. on Mobile Information Systems, Springer (2004) 121–134
2. Bettini, C., Mascetti, S., Wang, X.S., Jajodia, S.: Anonymity in Location-based Services: towards a General Framework. In: Proc. of the 8th Int. Conf. on Mobile Data Management (MDM), IEEE Computer Society (2007)
3. Samarati, P.: Protecting Respondents' Identities in Microdata Release. IEEE Trans. on Knowledge and Data Engineering **13**(6) (2001) 1010–1027
4. Sweeney, L.: k-Anonymity: a Model for Protecting Privacy. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. **10**(5) (2002) 557–570
5. Mokbel, M.F., Chow, C.Y., Aref, W.G.: The New Casper: Query Processing for Location Services without Compromising Privacy. In: Proc. of the 32nd Int. Conf. on Very Large Data Bases (VLDB), VLDB Endowment (2006) 763–774
6. Rich, E.: User Modeling via Stereotypes. Cognitive Science **3**(4) (1979) 329–354
7. Aggarwal, C.C.: On k-Anonymity and the Curse of Dimensionality. In: Proc. of the 31st Int. Conf. on Very Large Data Bases (VLDB), ACM (2005) 901–909