

# Context-Driven Prevention of Unintended Identity Disclosure

Davy Preuveneers, Liesje Demuynck\*, Bart Elen, Kristof Verslype, Bart De Decker, Yolande Berbers, and Pierre Verbaeten

Department of Computer Science, K.U.Leuven  
Celestijnenlaan 200A, B-3001 Leuven, Belgium,  
{davy, liesje, barte, kristov, bart, yolande, pv}@cs.kuleuven.be,  
<http://www.cs.kuleuven.be>

**Abstract.** In this paper we discuss the problem of privacy and information disclosure in ubiquitous and pervasive computing environments. We describe a framework that uses context information to selectively disclose personal information to service providers while keeping the human-computer interaction non-intrusive. For each interaction with a service provider, our framework will take previously disclosed personal information, user preferences and available privacy enhancing technologies into consideration to conduct an optimal information revealing strategy in a particular situation. The use of context-awareness in our framework provides the ability to infer revealing information and enables the user to adjust the privacy control in a non-intrusive manner.

## 1 Introduction

In the next decade, ubiquitous computing will gradually appear in the workplace, in public areas as well as in the home environment. This new computing paradigm promises to change drastically the human-computer interaction. Small embedded and networked computing systems provide 24/7 access to information and services, and this in a non-intrusive way [24].

A tourist's PDA will, in the arrival hall of the airport, gather information about hotels, taxis, public transportation, places of interest; it might also read private messages destined for its owner from the bulletin board or buy the local newspaper, and perhaps reserve a table in a restaurant nearby. It may need to identify its owner to the customs or immigration officer, or show a valid visa for the visited country. The thirsty tourist will have to prove his adulthood when ordering a beer in the snack bar. In all these interactions the user (or its representative, the PDA) needs to disclose personal data (name, age, address, etc.) to other people or service providers.

In a face-to-face transaction it is assumed that customers are better able to assess their benefit when disclosing personal information. Still, people have given away passwords to colleagues or used the same password for all kinds of

---

\* Research Assistant of the Research Foundation - Flanders (FWO - Vlaanderen)

services. In the ubiquitous computing setting, the number of interactions with personal handheld devices will increase considerably and the problem of privacy is worsened by the fact that ubiquitous computing will make it easier to collect personal information anytime and anywhere without any knowledge of the individual. Access control, personalization, authentication, accountability and liability, all require the collection of at least some personal data by the service provider. However, once this ‘personal data’ has been disclosed, it might live forever and be used anywhere. The owner of the data cannot control how the collector will use the data. This personal data can be collated, linked with transactional data, processed by powerful data mining techniques and assembled into user profiles that may become so detailed, that identification becomes possible. To avoid these troubling implications for an individual’s privacy, this paper presents a system architecture for privacy-sensitive ubiquitous computing focusing on anonymity and aiming to prevent unintended disclosure of personal data. The main characteristics of the framework are:

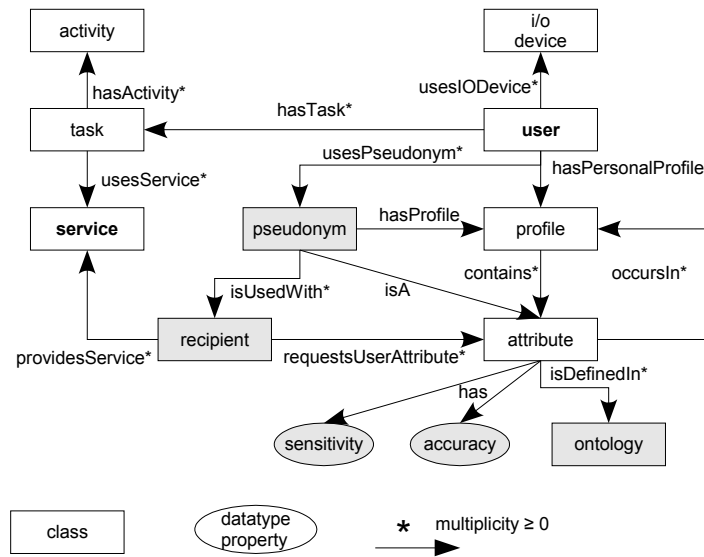
1. Decreasing the identifiability of user attributes before disclosure, i.e. reducing the level of detail in order to keep the number of matching candidates high.
2. Inferring implicitly known information based on previously disclosed user attributes and analyzing the impact of disclosing correlated user attributes.
3. Using context information to determine an appropriate information revealing strategy and to keep the human-computer interaction non-intrusive.

For situations where personal information exchange is desirable, our framework keeps track of which information has been disclosed before. Based on these previous interactions, the current context of the individual and the sensitivity of the requested user attributes, the framework will determine whether these user attributes can be disclosed to a known or unknown third party. This framework allows a PDA to protect its owner’s privacy by minimizing or eliminating the collection of personal data. In particular, the PDA will manage the individual’s personal information attributes and, given a particular context, send a subset of these attributes upon request using *Privacy Enhancing Technologies* (PET) with a proven track of record [1, 2, 5, 7].

In section 2 we describe how information is blurred and knowledge on the current context is used to selectively disclose user attributes. In section 3 we describe the privacy enhancing techniques that are used in our privacy framework. Section 4 elaborates on the building blocks of the framework and discusses the information disclosure strategy. In section 5 we evaluate an initial implementation of our framework. Section 6 provides an overview of related work. We end with conclusions and future work in section 7.

## 2 Context-aware information disclosure

There is a growing interest in context-awareness [9] for making mobile devices aware of their users and their physical environment. Context is often defined as any information that can be used to characterize the situation of an entity



**Fig. 1.** User ontology concepts for selective attribute disclosure

and often includes current location and time, user preferences and activities, available devices, services and resources in the vicinity. To support flexibility and mobility, context information is used to personalize services and to adapt services to heterogeneous networking environments and to a device’s capabilities [19]. The intelligent device is capable of acquiring and aggregating this context information, allowing context-aware applications to support user tasks by acting autonomously on their behalf. The use of context makes computing and communication transparent to the users in day to day activities, but the often hidden collection and combining of information has a high impact on a individual’s privacy. In the following subsections, we will model privacy aspects of contextual concepts and show how they are used when disclosing personal information to third parties.

## 2.1 Modeling user attributes for selective disclosure

The ease with which personal information can be gathered and the ability of context-aware systems to infer revealing information from loosely coupled data cause some serious implications on an individual’s privacy. As a countermeasure, our privacy control system uses our context infrastructure [19] to analyze the selective disclosure of sensitive user attributes before exchanging them with other parties in a particular situation. It is paramount to have an adequate modeling of currently and previously disclosed user attributes to understand the semantic meaning, the sensitivity, the relationships and correlations with other user attributes in a particular context. The history of information exchange al-

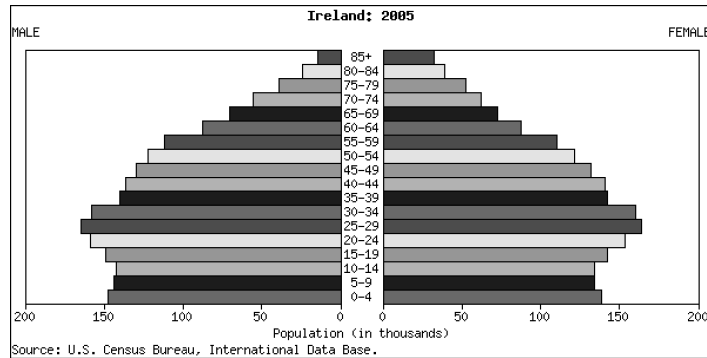
allows the context system to infer if other sensitive information can be derived by the collector. To represent this contextual information, we use our context ontology [8] and other domain specific ontologies. This context ontology defines concepts and relationships with respect to *Users*, *Platforms*, *Services* and the *Environment* to address context-driven adaptation of mobile services. See Figure 1 for an overview of the user ontology concepts in grey that were added to model the privacy aspects of user attributes.

A **user** is an individual whose privacy is to be protected. A user is characterized by a number of **attributes**, i.e. information with a certain privacy **sensitivity**. Privacy is ensured by keeping the number and the **accuracy** of the disclosed and linkable attributes low so that a person is not identifiable within a larger group of individuals. Information **blurring** is an attempt to obscure information and to reduce the sensitivity of user attributes by decreasing the accuracy as much as possible. This is easily achieved by generalizing user attributes into larger categories. For example, suppose the user does not wish to disclose his age, then the value of the *Age* attribute **age=27** can be blurred to the range **age>=18**, i.e. the category of adults. The **recipient** is the entity, e.g. a service provider, that requests information attributes from the user. A **user profile** is the set of attributes and other transactional data (selected services, bought items, accessed documents, etc.) revealed to a specific recipient. User profiles are **linkable** if they contain the same distinguishing (combination of) attributes. Such a linkable attribute is a **pseudonym**. It provides an anonymous identity and thus only identifies ‘a’ person and nothing else. A pseudonym is used to exchange attributes with a certain recipient and may be reused later on. The disclosed attributes are included in the corresponding profile of the pseudonym to keep track of what is known by the recipient. If a recipient finds an identifying set of personal attributes in two different pseudonym profiles, then he knows that both pseudonyms belong to the same user and the collected information can be combined. Each attribute is defined in an **ontology** that models the relationships with other attributes. These ontologies are used to infer previously undisclosed user attributes or to predict the sensitivity of disclosure. The **anonymity set** is the set of possible users fitting a set of attributes and must be kept as large as possible to avoid identification of the individual.

## 2.2 Sensitivity and blurring of user attributes before disclosure

Our framework focuses only on information being exchanged by the user or the personal handheld on his behalf. It does not consider the individual under surveillance with no knowledge of the personal attributes being sensed, or on third parties having side channels to find out about the individual. As such, the procedure to avoid unintended disclosure of identifiable information is solely based on the sensitivity of the disclosed attributes and the selective context-aware disclosure.

The **sensitivity**  $s$  of an attribute  $a$  with value  $v$  (with  $0 \leq s(a = v) \leq 1$ ), determines how identifiable the attribute is given its value. A rarely occurring attribute value decreases the uncertainty of identifying a particular individual



**Fig. 2.** Population pyramid of Ireland

more than a more common attribute value. For example, attributes with sensitivity 1.0 are typically identifying attributes (an individual’s name or social security number) and decrease the anonymity set with 100%, i.e. to a single individual. An attribute with sensitivity 0.5 (the user’s gender) reduces the size of the anonymity set on average with 50%. This sensitivity is modeled based on information gathered from statistical databases. One such statistical information source is the population pyramid of a particular country, as shown in Figure 2. The sensitivity of an attribute’s value can be reduced if an individual has the ability to blur information to a certain extent. Blurring is an enforced loss in **accuracy** to intentionally reduce the sensitivity of the attribute’s value.

### Blurring a single user attribute

Blurring numerical attributes is achieved by randomly selecting a value from the interval defined by introducing a relative or absolute error on the real value. For example, assuming the age of an individual is 40, then decreasing the accuracy to 90% results in the age  $A$  to be randomly chosen in the range:  $A \in \{ 36 .. 44 \}$ . Given the statistical information in Figure 2, the age of about 12% of the population complies with this interval, whereas the occurrence of 40 year old individuals is less than 2%. As such, randomly choosing the age in the 90% accuracy interval would reduce the sensitivity from 0.98 to 0.88. However, for the same 90% accuracy value, the age attribute of a child (**age=8**) is blurred less as it would be randomly chosen from a smaller interval:  $A \in \{ 7 .. 9 \}$  versus  $A \in \{ 36 .. 44 \}$ . Both relative and absolute errors need to be introduced to ensure that the occurrence of the values in the interval is large enough.

Non-numeric attributes, such as the current *Location*, can be blurred by generalizing the value into a larger category, for example, from `location=<address>` to `location=<city>` or `location=<country>`. Such attributes require extra knowledge defining the relationships that hold among these location concepts and the number of instances that match the same values. This hierarchy of concepts and relationships is represented using ontologies. This enables us to use the

generalization and inference capabilities provided by our context infrastructure. To summarize:

1. The sensitivity  $s$  of a disclosed attribute  $a$  is defined by the relative occurrence of its value  $v \in V = \{v_1..v_n\}$ .
2. The size  $n$  of  $V$  is determined by the generalization process of  $a$  to ensure that the relative occurrence of its value  $v$  is larger than a user defined threshold.

This technique is similar to the k-anonymity approach [23] but then in this case from the perspective of the user himself. However, under some circumstances it is not possible to blur information, for example, when the user is obliged to provide a digital certificate as proof of correctness or when a service will not work without the data. Therefore, information blurring should not be considered as the single panacea for this problem. Fixed rules to disclose certain user attributes under particular circumstances are needed in such cases. These rules will be discussed in a following subsection.

### **Disclosing multiple and possibly correlated user attributes**

Knowledge of the gender reduces the uncertainty of identifying a particular person on average with about 50%. However, this ratio can change considerably due to previously disclosed and correlated attributes. Given the prior knowledge that an individual has breast cancer, the gender becomes a very sensitive attribute as male breast cancers only account for approximately 1% of all breast cancer cases. On the other hand, because the male/female ratio is so skewed, not wanting to disclose the gender will rise the suspicion that the gender is male. For a male, it is safer to blur the attribute and say he has ‘*cancer*’ instead of ‘*breast cancer*’.

*The user should never disclose an attribute if it is correlated to an undisclosed attribute with a highly unbalanced individual partitioning that may lead to near identification in the future.*

All disclosed attributes are stored in the profile of a particular pseudonym. The user may have several profiles, each stating the following items:

- A user pseudonym
- The recipient(s)
- The history of previously disclosed user attributes
- The maximum size of the user attribute list
- The maximum user attribute sensitivity
- The minimum degree of anonymity
- Contextual applicability constraints based on location, time, recipient, etc.

The list of previously disclosed attributes should be limited in size to avoid a huge exposure of personal and transactional information after accidental disclosure and linking of an identifying attribute. To summarize:

1. Without disclosure of any attribute the initial uncertainty  $u$  of identifying an individual  $p$  is  $u(p) = 1.0$  (or 100%). This means that the recipient can only randomly guess for individual  $p$ .

```

<owl:Class rdf:ID="Pseudonym" />
<owl:Class rdf:ID="Recipient" />
<owl:Class rdf:ID="Attribute" />
<owl:Class rdf:ID="Profile" />

<owl:ObjectProperty rdf:ID="hasProfile">
  <rdfs:domain rdf:resource="#Pseudonym" />
  <rdfs:range rdf:resource="#Profile" />
  <rdfs:type rdf:resource="&owl;#FunctionalProperty" />
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="contains">
  <rdfs:domain rdf:resource="#Profile" />
  <rdfs:range rdf:resource="#Attribute" />
</owl:ObjectProperty>

<owl:DatatypeProperty rdf:ID="sensitivity">
  <rdfs:domain rdf:resource="#Attribute" />
  <rdfs:range rdf:resource="xsd:int" />
  <rdfs:type rdf:resource="&owl;#FunctionalProperty" />
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:ID="accuracy">
  <rdfs:domain rdf:resource="#Attribute" />
  <rdfs:range rdf:resource="xsd:int" />
  <rdfs:type rdf:resource="&owl;#FunctionalProperty" />
</owl:DatatypeProperty>

```

**Fig. 3.** Privacy concepts incorporated within the OWL context ontology

2. Attributes and correlated attributes with a high sensitivity or an unbalanced individual partitioning should be blurred or not disclosed at all.
3. For each disclosed attribute  $a_i$  with sensitivity  $s$  and previously disclosed attributes  $a_{1..i-1}$ , the uncertainty  $u$  of identifying an individual  $p$  decreases as follows:  $u_{new}(p) = u_{old}(p) * (1 - s(a_i|a_{1..i-1}))$ .
4. The size of the pseudonym profile should be restricted in order to avoid accidental exposure of a large amount of personal information at once.

The above steps only work reliably under the assumption that all the required sensitivity information is available.

### Modeling user attributes and occurrence statistics in OWL

The Semantic Web has provided us with the OWL [17] specification language for describing ontologies. Ontologies are considered as the ideal candidate for knowledge representation and processing. The context ontology in [8] and privacy concepts have both been implemented in OWL. An excerpt of the privacy concepts in this context ontology is shown in Figure 3. An instantiation of this ontology showing occurrence statistics of previously mentioned user attributes is given in Figure 4.

### 2.3 Context-aware disclosure of user attributes

Context-awareness is used in our framework to automatically adapt the privacy policy for disclosing user attributes according to different circumstances and sit-

<pre> &lt;Age rdf:ID="#38"&gt;   &lt;occurrence&gt;1.9&lt;/occurrence&gt; &lt;/Age&gt;  &lt;Intersection rdf:ID="#001"&gt;   &lt;attribute&gt;     &lt;Gender rdf:resource="#Male" /&gt;   &lt;/attribute&gt;   &lt;attribute&gt;     &lt;Age rdf:resource="#38" /&gt;   &lt;/attribute&gt;   &lt;occurrence&gt;0.9&lt;/occurrence&gt; &lt;/Intersection&gt;  &lt;Attribute rdf:ID="Age"&gt;   &lt;value&gt;38&lt;/value&gt;   &lt;accuracy&gt;90.00&lt;/accuracy&gt;   &lt;sensitivity&gt;88.00&lt;/sensitivity&gt; &lt;/Attribute&gt; </pre>	<pre> &lt;Gender rdf:ID="Male"&gt;   &lt;occurrence&gt;49.55&lt;/occurrence&gt; &lt;/Gender&gt;  &lt;PriorIntersection rdf:ID="#002"&gt;   &lt;prior&gt;     &lt;Cancer rdf:resource="#BreastCancer" /&gt;   &lt;/prior&gt;   &lt;attribute&gt;     &lt;Gender rdf:resource="#Male" /&gt;   &lt;/attribute&gt;   &lt;occurrence&gt;1.00&lt;/occurrence&gt; &lt;/PriorIntersection&gt;  &lt;Attribute rdf:ID="Gender"&gt;   &lt;value&gt;Male&lt;/value&gt;   &lt;accuracy&gt;100.00&lt;/accuracy&gt;   &lt;sensitivity&gt;50.45&lt;/sensitivity&gt; &lt;/Attribute&gt; </pre>
--	---

Fig. 4. OWL instantiation showing occurrence statistics of user attributes

```

<PrivacyPreference rdf:ID="pref001">
  <condition>
    <Recipient rdf:about="&office;#Secretary" />
  </condition>
  <allow>
    <Attribute rdf:resource="#Salary" />
  </allow>
</PrivacyPreference>

```

Fig. 5. OWL instantiation of a privacy preference

uations. A user does not want to be overwhelmed with a considerable amount of personal information requests, nor does he want to continuously change the privacy policy whenever the context changes. Therefore, context-constrained privacy preferences are part of the system and stored within each user profile as mentioned earlier. These context constraints can define conditional consents on the recipient, the current location and time, general or recipient specific user attributes, etc., to have access to sensitive attributes or categories of attributes. As such, privacy preferences keep the human-computer interaction non-intrusive. Examples of such contextual conditions in a simplified representation are:

- **if** *recipient*='Me' **allow all**;
- **if** *location*='Home' **allow all in** 'Personal Attributes';
- **if** *time*='9\_to\_5' **and** *recipient*='Boss' **allow** 'Location';
- **if** *recipient*='Secretary' **allow** 'Salary';

Each user profile can be distinguished by a particular user pseudonym. It specifies one or more constraints like the ones above and includes the user attributes which can be disclosed under these circumstances. Note that only one rule needs to match for a particular attribute to be disclosed. For example, the first two rules may both match in a specific context, but the first rule is the least restric-

tive. When none of the contextual constraints match, then the attribute request is either rejected or either allowed using a new profile with the current context as a conditional restriction upon user request. These context-aware privacy preferences are also converted to OWL for automated processing, of which a simple example is shown in Figure 5.

If the required sensitivity information is available, then the attributes can be blurred before disclosure to reduce the sensitivity. As such, privacy preferences can be used, combined with the blurring technique, to reveal attributes only to trustworthy recipients, but sane defaults are required nonetheless.

### 3 Providing guarantees with Privacy Enhancing Technologies

Up to now we have only discussed the user side of the privacy story. However, if a user wants to enjoy a particular service, the provider may want to be able to enforce access control on the basis of verifiable user attributes, for example, to restrict services to users older than 18. With the previously discussed techniques, he has no such guarantees. Also, a recipient may require that the true identity of a user can be revealed for accountability reasons. Therefore, to share user attributes anonymously with recipients, privacy enhancing techniques are required. Also, when a user's hand-held announces its presence on a recipient's network, it must not disclose any identifiable or linkable information. Therefore, an essential prerequisite for the successful execution of any privacy enhancing technology is the use of anonymous or unlinkable communication [3, 18]. Depending on the context of where these PETs are to be applied, some technologies might prove to be more appropriate than others:

#### **Pseudonym systems**

A pseudonym system is useful in order to build up a reputation with a recipient. A pseudonym consists of both a public and a secret key. The public key is the actual pseudonym, while the secret key is kept private by its owner to prevent the use of a stolen pseudonym. Later on, the user can access the service by showing his pseudonym and proving ownership of the corresponding secret key.

#### **Group signatures**

Group signatures [7] allow a user to sign statements in name of a given group. Anyone can check the validity of the signature and thus be assured that the signer indeed belongs to the group. However, nobody is able to identify the signer within the group. Group signatures can be used to blur the identity of a user. For example, the user can prove to a shop that he belongs to a discount-eligible group.

#### **Pseudonym certificates**

A pseudonym certificate is a binding between a public key and a list of personal user information, signed by a trusted third party. The public key can be

seen as a pseudonym, the corresponding secret key is used to prove ownership of the certificate. Pseudonym certificates allow users to reveal personal information while remaining unidentifiable. Different uses of the same certificate are linkable. Also, by proving ownership of the certificate, a user will automatically reveal all attributes incorporated into the certificate.

### **Anonymous credential systems**

Credential systems [5, 2, 1] allow for anonymous yet authenticated and accountable transactions between users and organizations. A user can pseudonymously obtain a non-transferable and revocable credential from an issuer. This credential can then be shown to a verifier, without different shows of the same credential being linkable to each other or to their issuing protocol. Furthermore, a credential may contain attributes which can be selectively disclosed to the verifier. Current credential systems use a combination of pseudonyms, blind signatures and zero-knowledge proofs. The latter are two-party interactive protocols between a prover and a verifier, such that after successful execution the verifier did not gain any new information apart from the validity of the shown statements. For example, a user can prove being older than 18 without actually disclosing his age.

Other privacy enhancing technologies [4, 6] can be used as well given a particular context.

## **4 Building blocks of the privacy framework**

The privacy framework is built on top of a context-awareness infrastructure [19] which is responsible for the gathering and storing of, reasoning on and dissemination of contextual information on the user and the services the user is interacting with. Our privacy infrastructure reduces the amount of personal information that would otherwise be disseminated unfiltered to recipients upon their request. See Figure 6 for an overview. The system itself has been built using a component based approach to provide a better QoS in a mobile computing environment [20].

### **4.1 Context Management: Source of information**

To determine the context of a user, the framework makes use of different information providers. These information providers include (1) *Sensors* for time and location tracking, etc., (2) *User Profiles* containing privacy preferences and transactional information input of previous interactions; and (3) *Third Parties* such as information retrieved from public databases.

Our context management system represents all concepts and user attributes as a knowledge base of ontologies in OWL. The use of these semantic models combined with the Jena 2 Ontology and Reasoning subsystem [13] enables the reasoning on user attributes to infer new facts. Any derived user attribute may have an impact on the identifiability of the user and therefore should be known in advance to conduct an optimal information disclosure strategy.

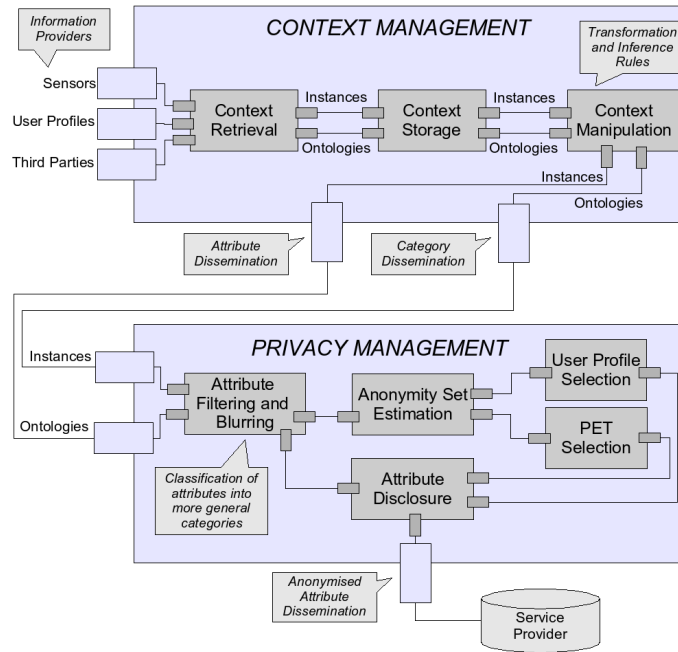


Fig. 6. Building blocks of the privacy preserving framework

## 4.2 Privacy Management: Anonymizing information

The privacy framework processes the information provided by the context management infrastructure before being disclosed to the recipient to ensure the privacy of the user. It blurs the list of attributes by using the ontologies to classify attributes into more general categories. For example, location information can be generalized from a detailed address to just the zip code or country.

When the attributes have been filtered and blurred, the *Anonymity Set Estimation* will investigate the sensitivity of the attributes and the overall identifiability of the user. Given the current context (time, location, etc.) it will collaborate with the *User Profile Selection* and the *PET Selection* components to decide which profile and privacy enhancing technique to use to make the user as anonymous as possible. For example, a pseudonym system may not provide the required proof of correctness for a particular user attribute whereas a pseudonym certificate may reveal more data than required or allowed by the privacy preference in place.

## 4.3 A non-intrusive information revealing strategy

The strategy of our framework is to provide a best effort to preserve the privacy of the user without overwhelming him with many questions whether certain

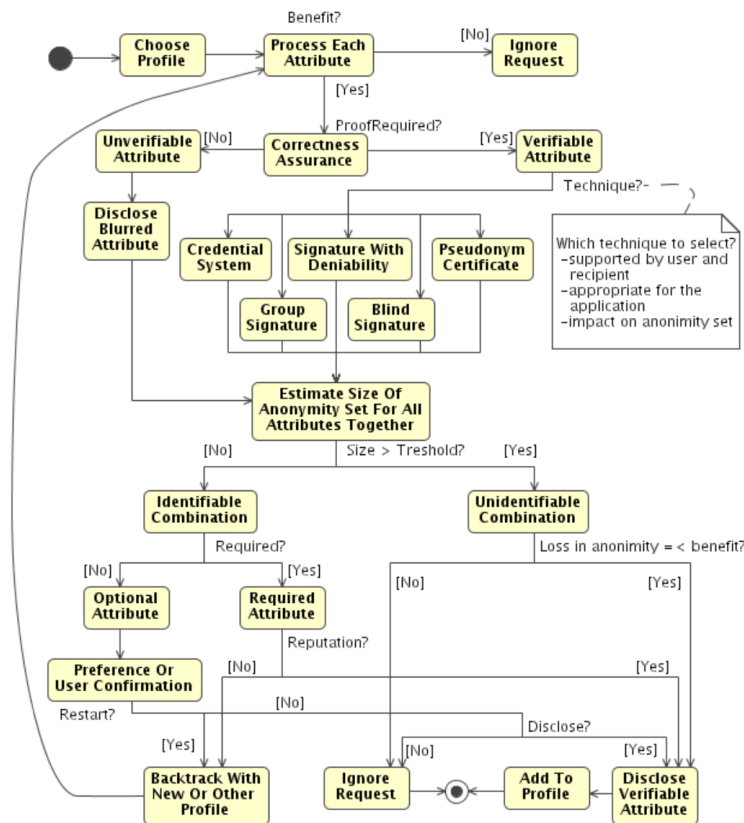


Fig. 7. Strategy of the privacy preserving framework

information may be disclosed to a given recipient. The main goal is to blur information and to make sure the information is unlinkable.

During the interaction, the recipient will request the user to submit attributes. He will specify the minimal accuracy, which attributes are required or optional, for which attributes a proof of correctness should be provided and the list of supported privacy enhancing techniques.

When the user receives a request from a recipient, the user has two choices. He will either wish to be anonymous and reuse the least revealing profile (or create a new one), or he will choose a specific profile to make use of a certain reputation built up with a recipient. Attributes previously revealed using the same profile can be linked by a pseudonym. Figure 7 demonstrates the course of our strategy which is largely implemented in the *Anonymity Set Estimation* component of the *Privacy Management* part in Figure 6. The first step is to check if a proof of correctness should be provided for the attribute. If not, the user submits a maximal blurred attribute. When correctness assurance is required,

an appropriate privacy enhancing technique should be selected. Which technique fits best, depends on which techniques are supported by both the user and the recipient and what the impact is on the anonymity set. Hence, the optimal technique can be case specific. For example, pseudonym certificates should be used as a last resort, as the user cannot blur the content of the certificate and cannot disclose only a subset of the attributes in the certificate.

Once a technique has been selected, the size of the anonymity set is estimated on the basis of all attributes to determine if disclosing the information would lead to an identifying combination of attributes. When the anonymity set is large enough, the attributes may be disclosed if the user profile has not become too large. If the anonymity set is too small, then the next step is to check if the attribute is optional or required. If the attribute is optional, then it is up to the user or his list of preferences to decide which step to take next. Otherwise, it depends on the user if he wishes to make use of his reputation and disclose the attribute with the current profile. If he does not want to rely on a reputation, the process starts all over with a new or existing profile for which a larger anonymity set is more likely. Typically, a user will not share a user profile between many recipients, but he may have several profiles with different degrees of anonymity for the same service provider.

## 5 Evaluation and implementation

The framework presented here is a work in progress and a thorough quantitative validation on practical real-life use cases is currently missing. However, we will reflect on the scenario mentioned in the introduction and describe how the proposed system would work there. We will also summarize the strengths and weaknesses of the framework and discuss some implementation details.

### 5.1 The airport scenario

For discovering information in the arrival hall support for anonymous communication in a network suffices. This is a pure technical issue that can be solved at the network layer of the communication protocol. Pseudonym registration with a travel agency allows to read private messages from the bulletin board, while a blind signature proves your newspaper subscription. Identifying yourself to the customs or immigration officer has a huge impact on the privacy. However, the options for the user are limited as he otherwise cannot enter the country. In this case, a pseudonym certificate can be an appropriate technique, while other partners can for accountability reasons rely on the customs to know the true identity of the person. In a snack bar users do not necessarily have to show their full identity. Proving adulthood can be managed by a credential system using a zero-knowledge proof that shows you are older than 18 and nothing else. As such, any privacy issue in this scenario can be fulfilled with the appropriate PET. The requirements from the user perspective are fulfilled as long as the sensitivity of the user attributes can be estimated accurately. Nonetheless, this is still an

open case as unknown dependencies between attributes might exist. However, we are convinced that due to the rather limited size of a user's profile and when having the proper privacy preferences in place, this issue can be overcome. The recipient requirements are fulfilled by appropriate privacy enhancing techniques that ensure information correctness.

## 5.2 Strengths and weaknesses

The strengths of our context-aware privacy control system depend on the presence of occurrence statistics for the required attributes. If not available, the sensitivity cannot be determined and must be declared by the user in its privacy preferences. The ability to infer other user attributes based on previously disclosed ones, depends on the presence, completeness and correctness of the ontologies describing the attributes. The more statistical and semantical information on an attribute is provided, the more accurate the sensitivity can be determined, the more computationally intensive the process becomes. Nonetheless, the proof-of-concept implementation is able to avoid unintended disclosures due to correlations in the user attributes that would otherwise be made blindly by the user.

## 5.3 Implementation details

The current infrastructure is implemented in the Java language and is backwards compatible with the J2ME Personal Profile so that it can be deployed on personal handheld devices running, for example, the IBM J9 virtual machine. The Java Cryptography Architecture (JCA) is used for the purposes of digital signing, verification of digital signatures, and handling digital certificates. The integration with a credential system is ongoing. For the OWL parsing and reasoning the Jena 2 framework of HP Labs is used. The current information representation in OWL reduces the complexity of managing data and enables the reasoning and processing of information in a uniform fashion. The framework has been tested to run on a Qtek 9090 handheld device, but the processing of OWL suffers from a performance penalty when using the most advanced OWL reasoner and the memory consumption in our prototype is considerable as all information resides in memory.

## 6 Related Work

Lederer et al. [16] proposed the 'face' metaphor as a permutation of privacy preferences. It is based on Fair Information Practices where a user can be notified of and give consent for the collection of personal information. To not overwhelm users with a large set of descriptive preferences, a small set of 'faces' is used to handle all personal information collections.

Langheinrich [15] introduced a privacy awareness system targeted at ubiquitous computing environments that relies on P3P for data collection. P3P (the

Platform for Privacy Preferences) specifies a machine-readable vocabulary that can be used by web-sites to communicate their data-management practices towards their visitors. This is done by sending, upon a HTTP-request from a visitor, a P3P policy file describing the capture, handling and use of personal information to this user. The visitor itself can then decide whether or not this policy is acceptable, using its own P3P preferences. This technique allows for the user to better control its personal data, provided an organization actually adheres to its privacy policies. As such, P3P is inherently a trust-driven approach to guarantee privacy as users need to trust web services that they conform to their privacy policies.

Credential systems are currently the most elaborated systems in which a user can control the degree of personal information revealed to a recipient. The concept was introduced by Chaum [5] and elaborated on by, among others, Camenisch et al. in *Idemix* [2] and Brands [1]. Their benefits are their wide applicability in privacy-sensitive applications and the possibility of the user to keep track of linkabilities between personal information. A drawback is their reliance on public key cryptography, which has its implications on performance.

Jiang and Landay [14] discuss the modeling of privacy control in context-aware systems and propose the use of information spaces to provide a way to organize information, resources and services around important privacy-relevant contextual factors.

Sweeney [23] proposes k-anonymity to provide privacy protection when sharing information on a large number of users with other parties in such a way that the users in the data set cannot be identified. Generalization and suppression is used to guarantee that each released record has at least (k-1) other records in the data set whose values are indistinct, i.e. to make re-identification more difficult.

Hong and Landay [12] have analyzed end-user and developer needs for privacy sensitive systems and have developed Confab, a toolkit that provides software architecture support and an extendible suite of mechanisms that application developers and end-users can use for managing privacy so that personal information is managed on a personal device as much as possible.

The PRIME [21] project (Privacy and Identity Management for Europe) has extensively studied Identity Management Systems and has recently proposed a general framework. In the terminology of this project, this paper describes a lightweight Identity Management Application (with emphasis on privacy and non-intrusiveness). The PRIME project also acknowledges that the core challenge of ambient intelligence environments will be addressing the balance between privacy and security as information about individuals will be gathered constantly and in an invisible way.

Several approaches for measuring and modeling anonymity have been proposed. Diaz et al. [10] describe a model to quantify the degree of anonymity in anonymous connection schemes under various types of attacks. Contrary to our system, the degree of anonymity in their model is independent of the number of users in the system.

Steinbrecher et al. [22] formalize the notion of anonymity and redefine definitions of anonymity in terms of linkability. The degree of linkability is based on the probability of items in a set being related. Finally, Van Herreweghen [11] provides a formal model allowing to measure the degree of anonymity in *Idemix*. This model is non-probabilistic and does not take into account the consequences of anonymous communication channels.

## 7 Conclusions and future work

In this paper we have discussed the privacy problem in a ubiquitous computing setting. We have identified several requirements from a user and recipient perspective and demonstrated their feasibility by means of small examples in a ubiquitous computing setting. We have proposed a novel non-intrusive information disclosing strategy based on the size of the anonymity set and the user profile while taking into account the context of information disclosure. The framework makes use of several proven privacy enhancing techniques that protect the privacy of the user and provide correctness assurance to the recipient if required.

Future work includes research on how to implement the right to erase disclosed attributes when they are no longer necessary. Other work will focus on refining the disclosing strategy when multiple attributes are stored in a pseudonym certificate instead of only one, as pseudonym certificates may reveal attributes not being asked for.

## References

1. Stefan A. Brands. *Rethinking Public Key Infrastructures and Digital Certificates: Building in Privacy*. MIT Press, Cambridge, MA, USA, 2000.
2. Jan Camenisch and Els Van Herreweghen. Design and implementation of the *idemix* anonymous credential system. In *ACM Conference on Computer and Communications Security*, pages 21–30, 2002.
3. David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. In *Communications of the ACM*, volume 24, pages 84–88, February 1981.
4. David Chaum. Blind signatures for untraceable payments. In *CRYPTO*, pages 199–203, 1982.
5. David Chaum. Security without identification: Transaction systems to make big brother obsolete. *Commun. ACM*, 28(10):1030–1044, 1985.
6. David Chaum and Hans Van Antwerpen. Undeniable signatures. In *CRYPTO*, pages 212–216, 1989.
7. David Chaum and Eugène van Heyst. Group signatures. In *EUROCRYPT*, pages 257–265, 1991.
8. Davy Preuveneers et al. Towards an extensible context ontology for ambient intelligence. In Panos Markopoulos, Berry Eggen, Emile Aarts, and James L. Crowley, editors, *Second European Symposium on Ambient Intelligence*, volume 3295 of *LNCS*, pages 148 – 159, Eindhoven, The Netherlands, Nov 8 – 11 2004. Springer.

9. Anind K. Dey. Understanding and using context. *Personal Ubiquitous Comput.*, 5(1):4–7, 2001.
10. Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In Roger Dingledine and Paul F. Syverson, editors, *Privacy Enhancing Technologies*, volume 2482 of *Lecture Notes in Computer Science*, pages 54–68. Springer, 2002.
11. Els Van Herreweghen. *Unidentifiability and Accountability in Electronic Transactions*. PhD thesis, Katholieke Universiteit Leuven, 2004.
12. Jason I. Hong and James A. Landay. An architecture for privacy-sensitive ubiquitous computing. In *MobiSys '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pages 177–189, New York, NY, USA, 2004. ACM Press.
13. HP Labs. Jena 2 - A Semantic Web Framework. <http://www.hpl.hp.com/semweb/jena2.htm>, 2004.
14. Xiaodong Jiang and James A. Landay. Modeling privacy control in context-aware systems. *IEEE Pervasive Computing*, 1(3):59–63, 2002.
15. Marc Langheinrich. A privacy awareness system for ubiquitous computing environments. In *UbiComp*, pages 237–245, 2002.
16. Scott Lederer, Anind K. Dey, and Jennifer Mankoff. Everyday privacy in ubiquitous computing environments. In *UbiComp 2002 Workshop on Socially-informed Design of Privacy-enhancing Solutions in Ubiquitous Computing*, 2002.
17. Deborah L. McGuinness and Frank van Harmelen. Owl web ontology language overview, w3c recommendation 10 february 2004, February 2004.
18. Andreas Pfitzmann and Michael Waidner. Networks without user observability: Design options. In *Advances in Cryptology- EUROCRYPT '85: Proceedings of a Workshop on the Theory and Application of Cryptographic Techniques*, pages 245–253, April 1985.
19. Davy Preuveneers and Yolande Berbers. Adaptive context management using a component-based approach. In Nancy Alonistioti and Lea Kutvonen, editors, *Proceedings of 5th IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS2005)*, Lecture Notes in Computer Science (LNCS), Athens/Greece, June 2005. Springer Verlag.
20. Davy Preuveneers and Yolande Berbers. Automated context-driven composition of pervasive services to alleviate non-functional concerns. In Ghita Kouadri Mostefaoui and Patrick Brezillon, editors, *Proceedings of the ICPS'05 International Workshop on Software Aspects of Context (IWSAC'05)*, pages 1–8, Santorini, Greece, July 2005. CEUR Workshop Proceedings, ISSN 1613-0073, online [CEUR-WS.org/Vol-150/paper4.pdf](http://CEUR-WS.org/Vol-150/paper4.pdf).
21. PRIME. Privacy and Identity Management for Europe. <http://www.prime-project.eu.org>, March 2004.
22. Sandra Steinbrecher and Stefan Köpsell. Modelling unlinkability. In Roger Dingledine, editor, *Privacy Enhancing Technologies*, volume 2760 of *Lecture Notes in Computer Science*, pages 32–47. Springer, 2003.
23. Latanya Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):571–588, 2002.
24. Mark Weiser. The Computer for the Twenty-First Century. *Scientific American*, pages 94–10, September 1991.