# Balancing smartness and privacy for the Ambient Intelligence

Harold van Heerde[1], Nicolas Anciaux[2], Ling Feng[1], and Peter M. G. Apers[1]

[1] Centre for Telematics and Information Technology, University of Twente, The Netherlands {h.j.w.vanheerde,ling,apers}@ewi.utwente.nl
[2] INRIA, France Nicolas.Anciaux@inria.fr

**Abstract.** Ambient Intelligence (AmI) will introduce large privacy risks. Stored context histories are vulnerable for unauthorized disclosure, thus unlimited storing of privacy-sensitive context data is not desirable from the privacy viewpoint. However, high quality and quantity of data enable smartness for the AmI, while less and coarse data benefit privacy. This raises a very important problem to the AmI, that is, how to balance the smartness and privacy requirements in an ambient world. In this article, we propose to give to donors the control over the life cycle of their context data, so that users themselves can balance their needs and wishes in terms of smartness and privacy.

## 1 Introduction

A smart, anticipating, and learning environment will have a great impact on privacy. Ambient Intelligence will be everywhere, is invisible, has powerful sensing capabilities, and most of all has a memory [1]. One of the main difficulties with privacy in the ubiquitous computing, is the way how data is collected. When making a transaction with a web shop, it could be quite clear which kind of data is exchanged. Ubiquitous computing techniques however, such as small sensors, active badges [2], or cameras equipped with powerful image recognizing algorithms, often collect data when people are not aware of it [3, 4]. In that case it is possible that people *think* they are in a closed private area (such as coffee rooms), but in *reality* they could be monitored by sensors in that room without being aware of it. This leads to asymetric information [3]. Xiaodong *et al* state that the presence of asymmetric information is the heart of the information privacy problem in ubiquitous computing. In environments with significant asymmetry between the information knowledge of *donor* and *collector*, negative side effects as privacy violations are much harder to overcome.

Several techniques have been proposed in the literature which let donors of the data specify privacy policies, in order to give control about their data to the owners of that data [5, 6]. Although such policies are rich enough to let people control who, when, how long, and what kind of information can be disclosed to specific applications, enforcing those policies is usually done through access control. Only relying on access control mechanisms to protect against

unauthorized disclosure of data, is not sufficient enough in terms of privacy protection [7, 8]. Perhaps the context databases can be trusted *now*, but they might not be in the future (due to the change of privacy regulation laws for example). Therefore, limited retention techniques are highly desirable to prevent large context histories to be disclosed.

A second problem of traditional privacy policies found in the literature is that they only provide means to express privacy wishes for specific applications. Usage of the data is known in advance, as is the purpose for which the data will be used. Purposes of traditional applications requiring (context) data are atomic in the sense that it is clearly known when purposes are fulfilled or not. For applications which will use context data to learn, infer, and thus to become smarter, it is not clear when such a purpose has been fulfilled, in other words, purposes are *non-atomic*. It is even unclear which services and applications will use the context data in the future.

For static databases containing large datasets with privacy sensitive data (like medical data), anonymization can be used to prevent disclosure of individual privacy sensitive data [9–11]. However, anonymization does not always give adequate privacy protection to everyone, and the usability of the data becomes sometimes lower than needed because individual privacy concerns and personal interests are not taken into account. Xiao *et al* [12] recognize this problem and propose to *personalize* the anonymization of privacy sensitive data.

The nature of data used in the ambient smart environments is different and more dynamic than that of traditional static data. The amount of smartness of applications is bound to the quantity and quality of the data they can use. The more accurate the data is, and the more data has been gathered from a certain individual, the better a smart application can learn from that data without user interaction [13]. The challenge is to find the best balance between the quality and quantity of data at the one side, and the privacy sensitivity of the data at the other side.

## 2 Motivation

Consider a working environment where employees can access the Internet and rank their visited websites. These Internet browsing behaviors are monitored and recorded in a database. Employees can query the database to find interesting websites based on the ranking, and discuss with other employees who have visited and ranked the websites. However, because the starting and end times when an employee made a website visit are also recorded, it is possible to deduce the duration that an employee spends on the Internet per day. Thus, most employees may not want the system to record such sensitive information in the database although the employees do benefit from the offered smart query services. To compromise the smartness and privacy requirements, a self-regulation of sensitive information could be like: one hour/day later, degrade the employee id to his/her group id or even faculty id. In this way, s/he can still use such a query service as

"*give me interesting websites visited by the people from the database group last week*".

## 3 Approach

We let people (the authors of monitored events) specify *Life-Cycle Policies* which will be bound to the acquired sensitive data [14]. Events are monitored and bound to a context tuple, which could contain *author*, *location*, *time et cetera*. This data is stored in a privacy aware context database system, which degrades the data progressively according to the policy. This way, context history can be considered as events (like *a door has been opened*) bound to attributes describing those events. The context values exhibit a certain level of accuracy based on domain generalization graphs. Such generalization graphs together form a n-dimensional space, in which each dimension represents the accuracy of an attribute of the original data tuple. We consider that those levels of accuracy can be classified given the privacy of the information they represent, such that all possible combinations of accuracies form a $n$-dimensional cube. A life-cycle policy can be viewed as a path specification in this cube. Triggered by events (we consider both time and contextual events), the accuracy of context tuples progressively decreases when specified conditions are satisfied.

## 4 Case Study

A prototype of a system which monitors the Internet browsing behavior of users has been implemented. Websites visited by users will be monitored, enabling smart services like ranking websites, contacting users of the same interests, finding interesting websites visited by members of a certain group, calculation of anonymized statistics, and so on. From the privacy perspective, collecting this data (including times when people were active on the Internet) makes it for malicious parties possible to deduce (for the user) confronting information. Users can specify their life-cycle policies (e.g., degrade time to hour and person id to group after one hour, degrade URL to category after one month, see Figure 1), which are attached with the data and will be stored and executed within a privacy aware context database. By specifying a life-cycle policy, users are sure that data which has been degraded can no longer be misused by malicious parties. Hence, the amount of smartness is reduced to decrease the possibility of misusing the data, consequently increasing privacy.

## 5 Acknowledgments

This work is funded by the Dutch organization for scientific research (NWO-Vidi project) and the Centre for Telematics and Information Technology of the University of Twente.

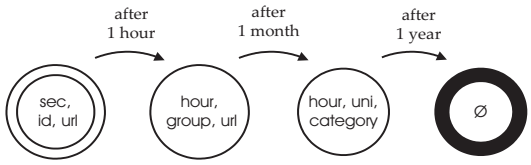


**Fig. 1.** *A LCP example with states (time, id, url) where $\phi$ stands for a deleted (or completely degraded) value*

## References


1. Langheinrich, M.: Privacy by design - principles of privacy-aware ubiquitous systems. In: UbiComp '01: Proceedings of the 3rd international conference on Ubiquitous Computing, London, UK, Springer-Verlag (2001) 273–291
2. Want, R., Hopper, A., Falcao, V., Gibbons, J.: The active badge location system. Technical Report 92.1, ORL, 24a Trumpington Street, Cambridge CB2 1QA (1992)
3. Jiang, X., Hong, J.I., Landay, J.A.: Approximate information flows: Socially-based modeling of privacy in ubiquitous computing. In: UbiComp '02: Proceedings of the 4th international conference on Ubiquitous Computing, London, UK, Springer-Verlag (2002) 176–193
4. Little, L., Briggs, P.: Tumult and turmoil: privacy in an ambient world. In: Workshop on Privacy, Trust and Identity Issues for Ambient Intelligence (Pervasive 2006). (2006)
5. W3C: Platform for privacy preferences (P3P) project. http://www.w3.org/P3P/ (2005)
6. Byun, J.W., Bertino, E.: Micro-views, or on how to protect privacy while enhancing data usability. Vision paper CERIAS Tech Report 2005-25, Center for Education and Research in Information Assurance and Security, West Lafayette, IN 47907-2086 (2005)
7. Agrawal, R., Kiernan, J., Srikant, R., Xu, Y.: Hippocratic databases. In: 28th Int'l Conf. on Very Large Databases (VLDB), Hong Kong. (2002)
8. Hong, J.I., Landay, J.A.: An architecture for privacy-sensitive ubiquitous computing. In: MobiSys '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services, New York, NY, USA, ACM Press (2004) 177–189
9. Sweeney, L.: k-anonymity: A model for protecting privacy. International Journal on Uncertainty Fuzziness and Knowledge-based Systems (2002) 557–570
10. Machanavajjhala, A., Gehrke, J., Kifer, D., Venkitasubramaniam, M.: l-diversity: Privacy beyond k-anonymity. In: CLDB. (2006)
11. Chawla, S., Dwork, C., McSherry, F., Smith, A., Wee, H.: Toward privacy in public databases. In: Theory of Cryptography Conference. (2005)
12. Xiao, X., Tao, Y.: Personalized privacy preservation. In: ACM Conference on Management of Data (SIGMOD). (2006)
13. Doom, C.: Get smart: How intelligent technology will enhance our world. Technical report, Computer Sciences Corporation: Leading Edge Forum (2001) A report available from www.csc.com.
14. Anciaux, N., van Heerde, H., Feng, L., Apers, P.: Implanting life-cycle privacy policies in a context database,. Technical Report TR-CTIT-06-03, University of Twente, P.O. Box 217 (2006)