

Personalization in Privacy-Aware Highly Dynamic Systems

Stefan Sackmann, Jens Strüker, Rafael Accorsi

Institute of Computer Science and Social Studies,
Department of Telematics, University of Freiburg, Germany

Fifteen years after Mark Weiser's inspiring paper on ubiquitous computing [¹], his vision has become technically feasible. Objects of everyday use are becoming increasingly interconnected and mobile communication of all bandwidth and devices of all sizes are used in various ways. Highly dynamic information systems (HDS) are emerging, bringing new challenges for the management of information systems: having to cope with *components* that enter and leave the system spontaneously and be *autonomous in their actions*. The changing and possibly conflicting requirements of the single components have to be taken into account leading to a *dynamic negotiation of requirements*. Moreover, such highly dynamic systems have to be able to contend with *constant growth* of communicated data avidly collected in various forms.

Solving the challenges of HDS is accompanied with a prospect of economic potential. A first realization is the present rollout of RFID by major retail groups worldwide. Currently, cost savings through process automation is of prime importance but the use of this technology in retailing goes beyond mere productivity improvements. Tagging items with RFID-chips in combination with other wireless technologies, equipping customers with mobile communication devices, and using upcoming sensor-networks allow, for example, personalizing services that have so far been successfully used in client-server e-commerce scenarios [²].

From Anonymous to Personalized Shopping Experience

Internet has substantially changed the way of personalization. As depicted in figure 1, three ways of tailoring services to customers can be distinguished. Firstly, online retailers use the Internet today on a large scale to recommend products to known customers according to their previous purchases or interests [³]. These *personalized services* build upon a one-to-one communication channel and require personal data as input factor. Secondly, retailers also use the Internet to offer *individualized services*, which do not require personal data. For instance, the recommendation of products according to the sequence of clicks, pages requested or items that have been added to the shopping cart. Since such individualized services can be realized without necessarily identifying the customers, they allow improved shopping experience, at the same time maintaining their anonymity at the same time. Thirdly, *universal services* such as a product search function or having a look at customer reviews need neither personal nor context data. Even so, they are a form of personalization because a single customer can choose a service that meets his needs at a particular time. All three kinds of services can be part of a personalization strategy with the objective of building up customer relationships, increasing customer satisfaction, generating a 'lock-in' situation, and in the end realizing higher turnover.

Today, consumers are faced with thousands of products in a physical store and have to walk far to find them. The introduction of highly dynamic systems in stationary retailing enables an electronic one-to-one communication channel and allows the collection of context data comparably cheaply and effectively as in current e-commerce environments. In grocery stores such as the 'Extra-Future-Store' in Germany, computers with a touch screen attached to a

¹ Weiser 1991.

² Murthi, Sarkar 2003.

³ Srikumar, Bhasker 2005.

shopping cart are deployed as personal shopping assistants (PSA) [4]. Today, these devices are equipped with a barcode reader and customers can interact with the retailer's information system over WLAN. Future forms of interaction may include customers using their mobile phones to communicate with RFID-tagged products and the retailer's information system [5]. Furthermore, sensors embedded in customers' clothing or products might also become the subject of interactions. Such a technical infrastructure enables the context of each customer to be taken into account, for example the current position within the store or the products in the cart. Combining all this context data in real time with customers' personal data and profiles already stored in the information system, the retailer can use the electronic interaction channels (PSA, mobile phone) to enrich customers' shopping experience.

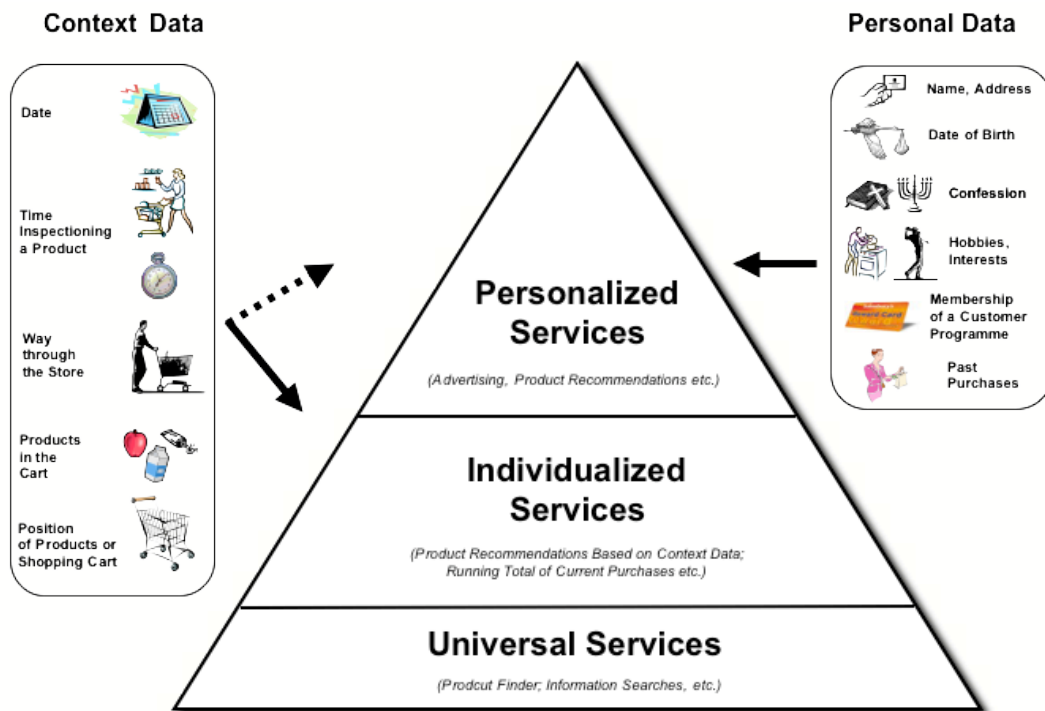


Figure 1: Personalization Pyramid

Imagine a customer equipped with an appropriate mobile communication device entering a store. To find a certain product, the customer can feed its name into the device and gets its location displayed. To obtain additional information, for example a list of possible recipes using this product or information about its origin, the customer scans the RFID-tagged article. Retailers are able to provide such *universal services* to all customers without necessarily taking the differences between each of them into account. *Individualized services*, however, additionally require data of the customer's context as input factor. For instance, a shopping list can be used to optimize the route through the store for time-sensitive or handicapped shoppers. Moreover, special offers or purchasing suggestions can be displayed according to the position of the cart within the store and the products in the cart. The mobile device can also show a running total of current purchases at any time, thereby enabling the customer to control expenditure. Finally, offering *personalized services* requires personal data such as name, age, purchasing history, or membership in a customer program. By identifying the customer, for example by means of an RFID-tagged customer card, the display can show further items as suggestions based on former purchases. Combining context and personal data

⁴ Litfin, Wolfram 2006.

⁵ Strueker, Sackmann 2004.

is also useful. On the way through the store, special offers can be displayed on the screen according to position *and* personal needs e.g. fat-free or whole food products. In this manner, allergy sufferers can, for instance, be warned about certain ingredients of products. Finally, thanks to personalized automatic checkouts the customer has no need to rummage for cash, pull out cards or have to queue [6].

Risks of Personalization

Although the economic potential of personalization in stationary retailing seems lucrative for retailers and customers, retail groups have slowed down their activities in this area. While Wal-Mart combined RFID-tagged articles with video surveillance, the German Metro Group tried to establish customer loyalty cards with embedded RFID tags [7]. However, after the sharp criticism of privacy activists, Metro decided to drop the use of RFID tags in cards and Wal-Mart also stopped their RFID-based surveillance (<http://www.bigbrotheraward.de/>). If customers were to refuse the processing of context data within the store in general, neither individualized nor personalized services would ever come into being. An analysis of the decisive privacy concerns shows that the loss of control over personal data worries customers. According to a survey of more than 1,000 U.S. consumers, two-thirds identified as a major concern the likelihood that RFID would lead to their data being shared with third parties [8].

User Acceptance Requires Transparency instead of Obscurity

Exploiting sensor networks, RFID identification, automatic video surveillance, localization technologies, and other technologies in HDS undermines the users' desire to control personal data. Extensive and unobservable data collection is an inherent characteristic of HDS:

- (a) Data is increasingly being collected without any indication. There will be no red indicator light on each device signaling the recording of data [9].
- (b) Data collection takes place without any pre-defined purpose, for example, the shopping cart continuously defines and reports its position to the retailers' information system. This information can be used for optimizing the store arrangement, for generating purchase suggestions as well as for identifying the customer.
- (c) Data once collected will be persistent and not deleted due to continuously decreasing cost of data storage.
- (d) Different devices record each event simultaneously from different view points, for example, a customer browsing a product is recognized by the smart shelf as well as by the video surveillance or the shopping cart. The combination of these different views allows, in combination with further context data, recognition or even identification of the customer.
- (e) Recording devices register multiple events simultaneously, for example, video surveillance can record customer *A* browsing a certain product, customer *B* passing the corridor, and customer *C* e.g. stealing a chocolate bar. The interpretation of the logged raw data for various purposes and the extraction of single events make the assignment of a valid privacy policy impossible.

The realization of a HDS leads to a paradigm shift of data collection and facilitates the relation of context data to individuals. The borderline between context data and personal data increasingly vanishes.

Today's Privacy Technologies Support Obscurity

⁶ Litfin, Wolfram 2005.

⁷ Chicago Sun-Times, Chipping away at your privacy, November 9, 2003.

⁸ RFID and Consumers: Understanding their Mindset, commissioned by Capgemini and the National Retail Federation and is available at http://www.nrf.com/download/NewRFID_NRF.pdf.

⁹ Langheinrich 2005.

The inherent data collection in HDS obliterates present-day privacy-enhancing technologies [10] because they are all based on concealing data - a privacy approach referred to as ‘obscurity’ throughout this paper. Today’s privacy mechanisms are incompatible with the objective of any retailer to provide both: personalization with useful services and assured privacy as well as security.

In table 1, a classification of privacy mechanisms is given. In the horizontal columns, the mechanisms are classified according to what they control: access or usage. While access control is usually understood as ex ante defined authentication and authorization, usage control extends access control and encompasses all those mechanisms that actually deal with the run time detection of privacy violations. In the vertical columns, guidelines, mechanisms and approaches for privacy are distinguished in whether they enable all three forms of personalization.

	Privacy Guidelines	Privacy Mechanisms	Current Examples	Enabling Personalization	Privacy Approach
Access Control	Controlled disclosure of data	Anonymity	MIX-Networks, JAP, Anonymizer [11]	No	Obscurity
		Pseudonyms/ Identities	Identity Management, e.g. IBM idemix, Liberty Alliance, Microsoft passport. iManager [12]	Yes	Obscurity
	Agreement on data collection	Policies, Seals, Certificates	P3P, EPAL, Privacy Aware System (PawS) [13]	Yes	Transparency based upon past
Usage Control	Transparent processing and usage	Monitoring processing of personal data	Obligations and Conditions	Yes	Transparency based upon past <i>and</i> the present
	Enforcing policy-compliance	Evidence Creation	Secure Logging and Auditing	Yes	Transparency based upon past <i>and</i> present with ex post enforcement

Table 1: Privacy and transparency

Anonymity, for example, prevents personalized services that require an identification of the customer. Pseudonyms and identity management, as the most favored solutions of science and industry, allow personalized services. Both privacy mechanisms follow the obscurity approach and rely on controlled disclosure of data, reducing such a disclosure to the minimal necessary to perform a given transaction. As a result, personalization is limited to the amount of disclosed data. However, the extensive and unobservable collection of context data for providing individualized services already allows the recognition of customers. This is because transactions are part of a chained process: e.g. filling the shopping cart, walking through the isles, scanning products and payment.

Obscurity, as a privacy approach for personalization in HDS, is inadequate. Once the access to data is granted, there is no control for customers as to how data is used – irrespective of the retailer’s initial intention. Proof of being an “honest” retailer acting according to data protection laws and the declared privacy policy can be produced by making data storage and data usage transparent. Different institutions providing a first step to transparency already exist: certification authorities, trusted third parties, privacy seals, code of conduct, or privacy

¹⁰ Langheinrich 2005.

¹¹ Köpsell, Wendolsky, and Federrath 2006.

¹² Wohlgemuth, Müller 2006.

¹³ Langheinrich 2005.

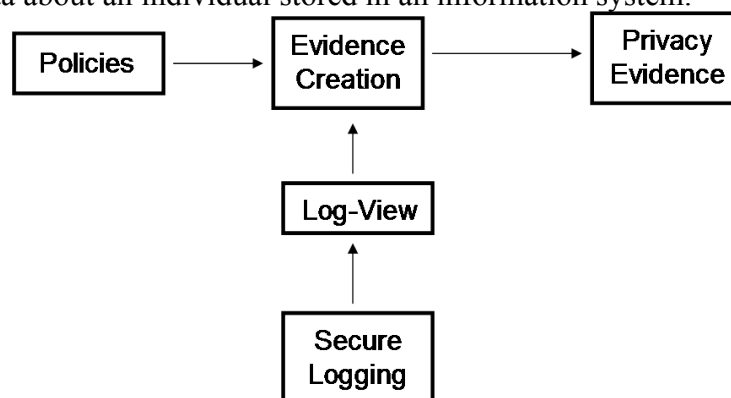
policies are implemented as a pre-defined agreement regarding the data usage. A promising approach is to supply tools to define individualized privacy and security policies and languages to express it. Currently, the most favored language for expressing privacy policies is P3P, the Platform for Privacy Preferences. P3P uses XML- specifications that state: (a) what kind of data is to be stored; (b) how data is to be used; and (c) its permanence and visibility, that is, how long data is to be stored and the corresponding access rights. Customers, admittedly, can express their desires but are not able to control the usage of their data. On the retailer's side, the rules for access are derived from the specified and possibly individualized privacy policies, for example by translating a valid P3P policy into EPAL (Enterprise Privacy Authorization Language), a formal language to express fine-grained enterprise privacy policies.

A highly dynamic system is only privacy-aware if it enforces formalized and personalized privacy policies. Such enforcement can be based upon past information (access control mechanisms), present and derived information (usage control). Enforcement can be achieved by an information system that has been proven to fulfill the desired properties, in particular self-limitation, and can expect to gain customers' trust by the resultant transparent access to personal data.

However, the characteristics of HDS restrict the effectiveness of formulated policies with regard to their adaptation. On the one hand, the autonomous components mean an increasing complexity for modeling the system and hinder the proof of their behavior. On the other hand, the changing manner of data collection rules out the assignment of a formulated privacy policy to personal data required for enforcing formulated policies: e.g., data collected outside the scope of a formulated policy, data collected by multiple devices is not integrated and related to a policy in real time, and data collected describing different events inherently interwoven may lead to conflicting policies. Technically, research could pursue the development of an adaptive 'P3P' or the control of the actual usage of data. First efforts try to prevent an unintended usage of data in real time as pursued, for example, by Park and Sandhu [¹⁴] or the article of Basin et al. in this issue.

Privacy Transparency by Evidence Creation

Instead of seeking for an *ex ante* approach to privacy transparency, in this paper we introduce the concept of *privacy evidence* for *ex post* enforcement of privacy policies. Transparency in HDS is provided by a cooperative mode between technology for detection and enforceable privacy contracts. The enforcement of privacy contracts requires for all involved parties the possibility to detect privacy violations – e.g. by means of audit – and document in a way that is acceptable as evidence, e.g., in a legal dispute. As depicted in Figure 2, the creation of evidence depends on: *policies* as reference for a compliant usage of data; and *log views* that encompass all data about an individual stored in an information system.



¹⁴ Park, Sandhu 2004.

Figure 2: Privacy Evidence Creation

Policies for Contract Representation

Today's state-of-the-art for contract representation is P3P. However, P3P cannot express composed privacy policies, in particular policies involving multiple, hierarchical departments or enterprises. These limitations are repaired by NAPS, the Novel Algebraic Privacy Specification [15]. Analogously to P3P, NAPS offers conjunction, composition and scoping operators for policies, but exhibit desirable algebraic properties. This extension is relevant in HDS, for it allows a distributed evaluation of composed policies. Although a practical realization is not yet available, NAPS demonstrates that there is no lack of expressive power regarding the representation of contracts. In other words, we can – at least theoretically – represent contracts adequately.

Secure Logging to Ensure Authenticity of Log Data

The second requisite for creating privacy evidence is log views generated from log data. However, standard logging mechanisms – such as syslog or syslog-ng – cannot be used for evidence creation, as they fail to ensure the necessary authenticity guarantees of log data. To provide such guarantees, secure logging is required and the central question concerns the characteristics log data must display to be accepted as evidence.

Authenticity of log data means: (a) *confidentiality*, i.e., log entries cannot be visualized or accessed by unauthorized individuals; (b) *integrity*, i.e., the log entries are accurate (entries have not been modified), complete (entries have not been deleted), and compact (entries have not been illegally added to the log file); and (c) *uniqueness*, that is, log data shall not allow for parallel realities. To realize these properties, proposals such as reliable syslog or Schneier/Kelsey's [16] are the only conceptual guidelines available today. Based on these existing guidelines, we develop a secure logging protocol to ensure authenticity of log data in a way suitable for evidence creation.

To realize this, standard cryptographic techniques are employed. Evolving cryptographic keys – hereby denoted by S – ensure not only confidentiality, but also forward integrity, i.e. if an attacker succeeds in taking over a logging device at time t , all the log data stored before t cannot be compromised. Hash chains, denoted by HC , guarantee integrity by creating interdependencies between entries. As a side effect, hash chains also provide tamper evidence and uniqueness guarantees of log data. Finally, entry-level access rights, denoted by AR , provide a way of controlling who has access to the log data. These access rights could be derived, e.g., from a user's privacy policies. Figure 3 illustrates how secure logging is realized, whereas its details and extension for remote collection of log data are found in [17].

¹⁵ Raub, Steinwandt 2006.

¹⁶ Schneier, Kelsey 1999.

¹⁷ Accorsi 2006.

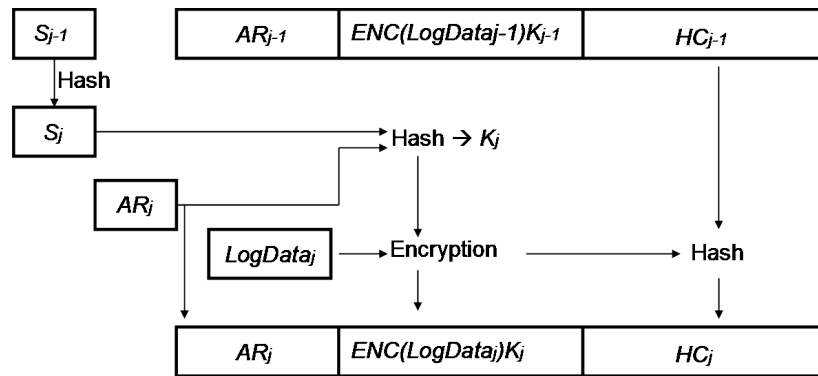


Figure 3: Realization of secure logging.

Log Views are a Basis for Evidence Creation

Albeit essential, secure logging is not enough to create evidence: views on logged data – conceptually similar to database views – are required but still not available. Log views are compilations of log entries encompassing all data collected about a user. In the case where log data can be directly assigned to a user and the related policies, generating log views can be tackled without further ado. For instance, in a P3P/EPAL setting, where the recorded data and the corresponding policies are stored together, a log view is just a query on log file parameterized by the user identification.

However, there are cases where this assignment is not directly possible or only within a certain degree of probability. In HDS there is a large variety of events which are recorded as isolated pieces of information without any relationship with the surrounding context. Moreover, HDS follow unspecified, unforeseen, and sometimes even chaotic patterns. This complicates the automated generation of precise log views. Techniques to generate log views include guessing particular situations and measuring their plausibility against known facts, as well as extensive data mining to search for specific patterns in recorded data. The results can, in some cases, doubtlessly be associated with the corresponding customer, but in the majority of cases a probabilistic estimation is the best one can get. Current efforts such as the Web Ontology Language provide an accurate description of context data and processes and could lead to more precise log views.

Limits of Log Views

To generate privacy evidence based on log view, its completeness remains an unresolved issue. In particular, it is currently impossible to exclude the existence of “shadow” log files hidden from a user. For example, “covert channels” within the system could redirect data to secondary log files not considered when generating log views. While trusted computing platforms could be used to attest the behavior of a data collector, guaranteeing completeness in HDS is even more challenging because of the inherent extensive and unobservable data collection. In consequence, regulatory institutions such as certification standards or legal advisory boards may be the only solution.

Conclusion

Highly dynamic systems enable several novel ways to personalize the relationship with the customer in stationary retailing. For this, the extensive collection and use of personal and context data are essential, but inherently raise privacy concerns: customers increasingly lose control over and awareness about which data is captured or how it is used. Surely enough, concerns of this kind considerably undermine the success of future personalization strategies. In highly dynamic systems, transparency with regard to the utilization of data is the only way to maintain privacy. The concept of *privacy evidence* we introduce in this paper is an initial

step in this direction, as it permits an objective view into the data collected about a customer. Evidence could be used as a “sword” for the customer to incriminate in the case of a misuse, or as a “shield” for the retailer to absolve in the case of a privacy-compliant usage. Privacy evidence paves not only the way to transparency, but also to an acceptable deployment of highly dynamic systems.

References

- Accorsi, R.: On the relationship of Privacy and Secure Remote Logging in Dynamic Systems. To appear in the Proceedings of the International Information Security Conference "Security and Privacy in Dynamic Environments", 2006.
- Köpsell, S.; Wendolsky, R.; Federath, H.: Revocable Anonymity. In Günter Müller (Ed.): ETRICS 2006, Lecture Notes in Computer Science, Volume 3995, pp. 208—222, Springer-Verlag, 2006.
- Langheinrich, M.: Personal Privacy in Ubiquitous Computing – Tools and System Support. PhD thesis No. 16100, ETH Zurich, Zurich, Switzerland, May 2005.
- Litfin, T.; Wolfram, G.: New Automated Checkout Systems, in „Retailing in the 21st Century: Current and Future Trends, Krafft, M. and Mantrala, Murali K. (Editors), p. 143 – 159, 2006.
- B. P. S. Murthi and S. Sarkar, "The Role of the Management Sciences in Research on Personalization," *Management Science*, pp. 1344-1362, vol. 49, no. 10, October 2003
- Park, J.; Sandhu, R.: The UCON_{ABC} Usage Control Model. *ACM Transactions on Information and System Security*, Vol. 7, No. 1, February 2004, Pages 128–174.
- Raub, D.; Steinwandt, R.: An Algebra for Enterprise Privacy Policies Closed Under Composition and Conjunction. In Günter Müller (Ed.): ETRICS 2006, Lecture Notes in Computer Science, Volume 3995, pp. 132—146, Springer-Verlag, 2006.
- Schneier, B.; Kelsey, J.: Security audit logs to support computer forensics. *ACM Transactions on Information and System Security*, 2(2):159–176, May 1999.
- Srikumar, K., Bhasker, B., Personalised recommendations in e-commerce, *Int. J. Electronic Business*, Vol. 3, No. 1, 2005.
- Strüker, J.; Sackmann, S.: New Forms of Customer Communication: Concepts and Pilot Projects. In: Proceedings of the Americas Conference on Information Systems (AMCIS '04), 6.-8. August, New York, USA 2004.
- Mark Weiser, "The Computer for the Twenty-First Century", *Scientific American*, pp. 94-10, September 1991.
- Sven Wohlgemuth and Günter Müller. Privacy with Delegation of Rights by Identity Management. In Günter Müller (Ed.): ETRICS 2006, Lecture Notes in Computer Science Volume 3995, pp. 177-191, Springer-Verlag, 2006.