Temporal Blurring: A Privacy Model for OMS Users

Rosa A. Alarcón, Luis A. Guerrero, and José A. Pino

Department of Computer Science, Universidad de Chile,
Blanco Encalada 2120, Santiago 6511224, Chile
{ralarcon, luguerre, jpino}@dcc.uchile.cl

Abstract. Stereotypes and clustering are some techniques for creating user models from user behavior. Yet, they possess important risks as users actions could be misinterpreted or users could be associated with undesirable profiles. It could be worst if users' actions, beliefs, and comments are long term stored such as in Organizational Memory Systems (OMS) where users' contributions are available to the whole organization. We propose a privacy model based on four privacy roles that allow users to control the disclosure of their personal data and, when recovered, blurs such data as time passes.

1 Introduction

User modeling addresses the need to improve user/computing system interaction. The system must have certain knowledge about users' preferences, strengths, weaknesses or other aspects relevant to the interaction to achieve this goal [7]. Two techniques for obtaining user information are stereotyping and clustering. Stereotyping captures pre-defined, default information about groups of people, while clustering techniques dynamically derive clusters of people with similar behavior under certain circumstances. Both techniques let the system predict user behavior, preferences or intentions. Thus, the software can adapt to the user and improve the interaction [8].

A promising area where users' models can help users to find the appropriate information is Organizational Memory (OM). An OM can be seen as the knowledge accrued by an organization and the set of mechanisms to preserve, distribute and reuse it [10]. An OM is immersed in an organizational setting and present important challenges regarding users privacy. An OM records users' actions, opinions, comments, etc. for long periods of time and make such data available to the whole organization. It is quite easy for an organization to derive user models not only from users' behavior but also from their knowledge, explicitly or implicitly stated in an OM. Some of these uses may be legitimate but others may be unethical and undesirable. In addition, people's knowledge, opinion and behavior change with time.

Some approaches have been proposed to support users' privacy ranging from anonymity to disclosure of user identity and are applied according to a privacy policy [6]. However, for OM users it is hard if not impossible to anticipate all future cases where such data could be retrieved and how a privacy policy will be applied.

In addition, OM content is created by collaborative interaction among colleagues. However, group members need information about others' status and actions. This

tradeoff between the need of parallel work visibility and privacy is well known [1], but surprisingly, it is neglected in most development proposals. If users have serious concerns about undesirable use of personal information (e.g. ideas, opinions) it is possible they refrain from making honest contributions to the OM, sustain another opinion outside the OM, or try to cover their identity by performing a false behavior. This undesirable situation will lead to the failure of the OM system.

2   Organizational Memory

Organizational memory (OM) refers to the stored information that can be reused for present decisions [10]. It allows capturing, organizing, disseminating and re-using the organization employees' formal and informal knowledge. OM content can be derived from individuals, organizational culture, transformation mechanisms, organizational structure, ecology and external information [10].

To allow users making sense of retrieved information from the OM it is important to present the context where it was created. Then, an OM system must also capture, store and distribute context-dependent knowledge. For instance, if a *lesson* is learned, then the *task* that *caused* the *lesson* and the relationship among both knowledge pieces is stored (context); when the *lesson* is retrieved, the *task* is also available, and vice versa, when the *task* is retrieved, the *lesson* is also available.

3   Privacy Strategies for Single and Groupware Users

OMS require particular privacy strategies due to three main reasons: a) the possibility of obtaining implicit knowledge, since it is unethical to attribute authoring of implicit knowledge; b) knowledge could be interpreted out of context, specially for subjective information such as evaluations; and c) the long-term nature of stored knowledge.

In groupware, privacy contradicts the need of sharing information about others (awareness). Awareness is crucial for workgroups' success because it allows efficient coordination and makes possible users be accountable for their actions and decisions. Bad privacy policies could hinder the interaction. Thus, ad hoc strategies are proposed or privacy is neglected in most research. Some strategies are: forbid access for non-members (secrecy); outgoing data filters, where users choose if an object is public, but they lose control once published [9]; social conventions, where users agree on common policies; social translucence, where actions visibility is based on reciprocity [3], and anonymization based on data distortion, e.g., aggregation, minimizing [9].

In single user systems, privacy is related to protect personal data such as name or credit card number of an identifiable person. Main concerns are the storage, transfer, unsought collection and processing of personal data; and its transfer to places with other privacy laws. Kobsa [6] proposes a reference architecture for pseudonymity in user-adaptive systems and mentions strategies ranging from secrecy to levels of anonymity such as: super-identification (authentication), identification (login), pseudonymity (users adopt a unique, linkable, unlinkable, unobservable or unidentifiable pseudonymous) and full anonymity (user cannot be identified).

3.1 Users' Privacy Roles

Experience with an OM system made us aware of two periods: 1) knowledge creation during intensive periods of collaboration, and 2) created knowledge is retrieved afterwards, with decreasing interest for authorship unless users are trying to find experts. Our aim is to find a strategy to support both collaboration and privacy.

We can identify at least four users' privacy roles or levels of identity disclosure: no privacy, alias-based identification, pseudonymity and anonymity in decreasing order. Table 1 shows the support those generic roles provide for collaboration. We consider meta-roles or users' stereotypes regarding privacy. They can also be applied in conjunction with other users' models (e.g. users' intentions when searching, task stereotypes such as "coordinator"), in a way guaranteeing users' privacy prevails. These roles are not contained in one another and neither can be arranged in a hierarchy, as we can observe from the properties in Table 1. Besides, it is desirable users could choose the kind of privacy role to be applied for certain circumstances.

Table 1. Impact of privacy roles on collaboration

User's Privacy role	Privacy Level	Collaboration Support
1. No privacy	None	Very high
2. Aliases	Low	High
3. Pseudonymity	Medium	Low
4. Anonymity	Very high	Very Low

4 Proposed Privacy Model

Our OM system is composed of a groupware subsystem capturing information while users work (PRIME) [4], and an Information Retrieval subsystem performing knowledge recovery and context retrieval implementing our privacy policy (OMUSISCO) [5]. PRIME (PRe-meetings Information Management Engine) is a Web-based system supporting a *collaborative* activity: asynchronous meetings preparation.

4.1 Privacy Strategy

If all or most PRIME users choose an anonymous profile, then their collaboration gets less effective; e.g., it would not be possible to know who was responsible for a task, or users could do free riding. However, anonymity is useful in collaborative systems because it allows users to participate in conversations or voting-systems without fear of reprisal [2]. If people use aliases, then it will be possible to identify poor contributors and perhaps motivate them. Of course, users could have more than one alias (otherwise they could be easily identified) and then a problem arises: user accountability and reward would be very difficult to achieve. Finally, users could be supported by pre-defined roles with various restrictions [6] and disclosure levels (e.g. coordinator), so they can have some control for protecting their identity.

Although each approach seems promising, none of them fully answers OM needs. An OM must gather information from users to reuse it in the future, but under this scenario, users could restrain of making sincere contributions because of fear of later stereotyping, misunderstanding and reprisal. Our model applies previous techniques and takes into account the passing of time by means of a progressive forgetting or *authorship blurring* function. This function is a metaphor of the real world: when remembering a conversation held some time ago one typically reminds "someone" said "something" but can not fully recall the author's name perhaps because the focus is on the subject of the conversation and not on the author's identity.

In our approach, users can choose to log into the system or make a contribution with any of the four privacy roles defined: 1) full identification (an organization account), 2) an alias name, 3) a role name and 4) anonymity. A name for each role is assigned to each user: e.g. *john@uchile.cl*, "*Doomsday*", "*Tester*" and "*Anonymous*" respectively. At the beginning, a first time frame (t_1) for a subset of participants (p_1, p_2, ... , p_k), a "*No privacy*" role is defined by default. Retrieved information related to such time frame and participants will show their full identity. However, users can choose any other role explicitly. For instance, in Fig. 1, during *time frame 1*, a user may choose to vote using his/her anonymous role. OMUSISCO will keep the user's choice: it will show "*Anonymous*" as the contribution author.

This approach makes possible to fully support users' needs for awareness information during an intense collaboration phase. After this phase (suppose it lasted one month), the discussion is closed (no further modifications are possible) and a new time frame is defined (t_2). OMUSISCO will blur authorship for information modified or created during *time frame 1* by replacing full identification with the corresponding alias name. Again, a user can choose another role when making a contribution. For instance, in *time frame 1* a user chooses the role "*No privacy*" explicitly and then creates an *argument* describing a paper written by him/her. Future retrieval of this information will always show the user's full identity (e.g. "john@uchile.cl" in Fig. 1).
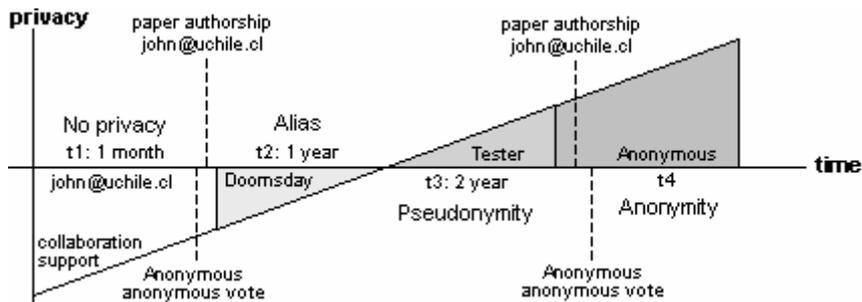

Fig. 1. Our approach progressively applies four privacy *roles* to a *subset* of *participants*

Under this scenario, there is no collaboration as the discussion is finished. Consider staff members retrieving knowledge from discussion of the previous project in this period: they could be interested in the topics and only referentially in the authors.

After a year, a third time frame is created and again, OMUSICO will distort author's full identification and alias name providing only the user role name (e.g. "Tester"). Finally, a fourth period is defined (e.g. after 2 years) where authorship will be blurred and regarded as "Anonymous" unless the author explicitly chooses another role such as in the "paper authorship" and "anonymous vote" events shown in Figure 1.

5   Conclusions

Techniques for retrieving and mining information make possible to discover otherwise unknown information. They help to find out user patterns of behavior, goals and needs, knowledge, so accurate user models can be derived. However, users may be concerned about possible unethical use of such information [6, 7, 8], and refrain to behave sincerely or fake their behavior. This concern may occur in organizational environments, but also includes open settings such as the Web. Poor user models can be derived if users distort their behavior due to perceived lack of privacy.

We grounded our model in an organizational setting such as the OM systems. Our privacy model is based on assumptions about users' privacy needs in such system. Such needs had been identified from the literature as well as from our experience. Our approach changes the system behavior in time, according to the users' privacy roles. The privacy roles encapsulate and describe characteristics of OM users regarding information privacy. Naturally a system implementing our model must guarantee the model itself is applied. We implemented our privacy model as part of the retrieval engine (OMUSISCO) of our OM system called PRIME. PRIME has been developed and initially tested with users at a large organization; the results are encouraging [4]. An OMUSISCO prototype has also been developed but not tested yet.

Acknowledgments

References

1. Borges, M.R.S., Pino, J.A.: Requirements for Shared Memory in CSCW Applications. Proc. of 10th Workshop on IT and Systems (WITS'00), Brisbane, Australia (2000) 211-216
2. Briggs, R.O., de Vreede, G.: Meetings of the Future: Enhancing Group Collaboration with Group Support Systems. Creativity and Innovation Management, 6(2) (1997) 106-116
3. Erickson, T., Kellogg, W.A.: Social Translucence: An Approach to Designing Systems that Support Social Processes. Transactions on Computer-Human Interaction, 7(1) (2000) 59-83
4. Guerrero, L.A., Pino, J.A.: Preparing Decision Meetings at a Large Organization. Proceedings of DSIage' 02, Cork, Ireland, Oak Tree Press (2002) 85-95

5. Guerrero, L.A., Pino, J.A.: Understanding Organizational Memory. Proceedings of 21st International Conference of the Chilean Computer Science Society, Punta Arenas, Chile, November, IEEE CS Press (2001) 124-132

6. Kobsa, A., Schreck, J.: Privacy through Pseudonymity in User-adaptive Systems. ACM Transactions on Internet Technology, 3(2) (2003) 149-183

7. Kobsa, A.: Supporting User Interfaces for All Through User Modeling. Proc. of the 6th Int. Conf. on Human-Computer Interaction, HCI, Yokohama, Japan (1995) 155-157

8. Schwab, I., Kobsa, A.: Adaptivity through Unobstrusive Learning. KI-3 (2002) 5-9

9. Sohlenkamp, M.: Supporting Group Awareness in Multi-user Environment through Perceptualization, Dissertation, Paderborn. GMD Research Series, No. 6 (1999)

10. Walsh, J.P, Ungson, G.R.: Organizational Memory. Academy of Management Review 16(1) (1991) 57-59