Explanations and Privacy in Intelligent Social Awareness Applications

Jörg Cassens, Anders Kofod-Petersen,
Sobah Abbas Petersen, and Monica Divitini

Department of Computer and Information Science,
Norwegian University of Science and Technology,
7491 Trondheim, Norway
{cassens|anderpe|sap|divitini}@idi.ntnu.no,
http://www.idi.ntnu.no/

Abstract. Explanations play an important part in the interaction with any intelligent system. This is particular important in context-aware and social awareness systems that regularly assume responsibility for a user and act proactively. Explanations are often generated using all available information. However, privacy issues in context-aware systems might dictate a limited distribution of information. The work presented here demonstrates how personal awareness-systems can fulfil different goals a user can have towards explanations, yet maintain a sensible level of privacy.

1 Introduction

Communication technology is often used to facilitate the exchange of information which gives the social networks of a person insight into what the person is doing and what his plans and expectations are. Examples for this are the twitter nanoblog service, www.twitter.com, status updates of facebook profiles, www.facebook.com, and the user status of instant messaging and voice over ip applications such as Google talk, talk.google.com or skype, www.skype.com.

These different means of conveying information about ones status are often decoupled from each other, although Web 2.0 mesh technologies like rss feeds and open web service api's can mitigate this to some extent. Second, control over who can receive this information is often not very fine grained. Status messages on IM clients are visible to all members of a user's contact list, and every finer grained distinction is a process of distributed social sense making (for example when the "busy" status means the user is interruptible by close friends, but not others). As a third problem, use of such systems is mostly restricted to traditional means of human-computer interaction, with the use of text message clients on mobile phones being only a partly exemption.

The ASTRA project [1] is addressing these issues by researching and implementing an architecture and end user tools for community oriented awareness applications using pervasive computing devices supplemented by traditional interfaces. In addition, we are working on interfacing our technology with existing

services, for example by automatically updating the twitter feed of the user, or integrating with existing instant messaging and voip protocols to initiate direct communication.

A typical usage scenario for ASTRA would be Alice who is writing a conference paper at her office and wants to go for a walk in order to take a break. She can use an ASTRA enabled device, for example a cube with motion and orientation sensors, to express her wish for a walk by giving this cube a specific orientation. This information is then made available to a pre-defined community of users, whose members can individually define how this information is presented. Bob has chosen to see updates of Alice's states on his picture frame, showing a picture of Alice and a specific colour. Bob can then decide whether he wants to contact Alice for a walk. For the time being, both publishing of awareness states and subscription of other person's states are defined via user defined rule sets. In our example, Alice has coupled the orientation of her cube with the wish for walk, and Bob has defined in his rule set that his subscription of Alice's wish for a walk should be displayed on his picture frame.

In human to human interaction, the ability to explain ones own behaviour and course of action is a prerequisite for a meaningful interchange, therefore a truly intelligent system should provide comparable capabilities. The shift from passive systems, to what could be regarded as a partnership between humans and intelligent artefacts, fosters the need for social adept system [2], in such a way that intelligent systems have to show certain abilities traditionally ascribed to humans [3]. Among these abilities we would count a system's ability to explain its behaviour. In order to make sure that the system can sufficiently explain itself, the following issues have to be addressed:

- The expectations users can have towards the explanatory capabilities of the system have to be analysed.
- The design of the system should make it possible to match these expectations, both in terms of
  - the knowledge model of the system (to make sure the system has sufficient knowledge to explain itself and its interactions with the user), and
  - the user interfaces (to make sure that explanations can be delivered to the user in a meaningful way).
- Explanations should cover both
  - capabilities to explain the state of the system itself (for example to be able to deliver meaningful and understandable failure diagnosis), and
  - capabilities to explain the interaction of the system (for example in case of a functioning system why certain information about other users is available or not).

In addition to these general issues, in the scope of an awareness project like ASTRA, we have to make sure that additional aspects of privacy are considered. While explanations about the inner working and the status of the system itself can be delivered without taking other users into account, we have to make sure

that explanations involving other users' published awareness states do not involuntarily disclose information but still satisfy the explanation needs of the user requesting that information as much as possible.

Another important aspect is that proactive systems that sense the world and are accessible primarily through behavioural interfaces should be able to explain their behaviour through these interfaces as well. We will not focus on this aspect in the course of this paper.

The paper is organised as follows: in the following section, we will introduce some basic concepts of awareness applications and describe the basic model underlying the ASTRA project. In Section 3, we will discuss some aspects of privacy in ubiquitous computing, and outline different models of privacy. This is followed by Section 4 which deals with explanations in social awareness applications. We show that some explanations are available which support user goals with explanation we have introduced earlier. We will also highlight the relation between explanation and privacy issues. We will conclude with a section discussing our results and sketching out future lines of research.

2 Awareness Applications

Awareness systems in the sense of this paper are systems that help users to build sustainable mental models about activities and communication wishes of other users. The theoretical background for ASTRA awareness applications is the focus-nimbus model, originally described by Benford et al. [4]. The authors use room metaphors as the basis for a spatial model to support communication between participants in virtual rooms. The basic idea is that people can not only visit different virtual rooms, but they can move around in these different rooms, and the (modeled) spatial characteristics of the rooms mediate the communication between different persons in the room. Two concepts are introduced; the *focus* represents a space in the room where a person targets his attention. People are more aware of objects in the focus then of objects outside. The *nimbus* is the counterpart, representing where the person locates himself in the room. Objects are more aware of a person if the object is located in the person's nimbus than when it is located outside [4, p. 220]. *Awareness* is defined through the interaction of focus and nimbus, and can be mathematically expressed through the spatial relation of a focus and a nimbus.

This model has been generalised by Rodden [5]. He extends the notions of focus and nimbus towards application areas without an explicit notion of spatial relations. Basically, he introduces a graph model for a domain, and awareness becomes a property of this graph. In its simplest form, the awareness measure is the length of the path between two users. Metaxas and Markopoulos have later presented a formal model which concentrates on the communication aspects of the focus-nimbus model [6]. Their model addresses issues of privacy by allowing for plausible deniability and deception.

In the terms of our example, Alice would make her wish for walk available by placing it on her nimbus. She can control to which community of users she pub-

lishes this aspect. Bob, on the other hand, would have a focus on this particular aspect of Alice's information.

3 Privacy in Awareness Applications

With regard to privacy in context-aware systems, Langheinrich [7] describes four properties highlighting privacy issues in ubiquitous computing: *ubiquity*, computers are everywhere; *invisibility*, computers disappear from the scene; *sensing*, sensors are becoming more precise; and *memory amplification*, storing of large amount of (sensed) data.

Many of the privacy issues in context-aware systems are related to the issue of *mutual awareness*. One part of this problem is about *disembodiment* and *dissociation*. When we encounter people in the real world we can receive information in many ways, such as position, voice level, facial expression and direction of gaze. In ubiquitous environments these communication channels are likely to be less effective. In real life people are guided by an intuitive principle: if you cannot see me, I cannot see you. Due to the potentially large number of sensors in an ubiquitous environment, this is not always true. Users may not always know exactly what information they are conveying, in what form, whether it is permanent, and to whom it is sent [8].

Jiang et al. [9] discuss the *principle of minimum asymmetry* when dealing with privacy issues in ubiquitous computing. This principle goes a long way towards handling the apparent asymmetric relationship between sender and receiver of information, as described by Bellotti et al. [8]. Jiang et al. argue that a privacy-aware system should minimise the asymmetry of information between data owners and data users. For an example, if a user does not wish to share his location he cannot expect others to share their location with him (regardless of their wish). The main principle of Jiang et al. is that [9, p. 7] (original emphasis):

> A privacy aware system should minimise the asymmetry of information between **data owners** and **data collectors and data users**, by:
> - **Decreasing** the flow of information from data owners to data collectors and users
> - **Increasing** the flow of information from data collectors and users back to data owners

Lederer et al. [10] argue that feedback and control is "the designer's opportunity to empower those processes (understanding and action), and they are the user's opportunity to practice them." The authors exemplify pitfalls in design of systems maintaining privacy using their personal experience. The pitfalls can be divided into two main groups: feedback and control. The feedback pitfalls are: *obscuring potential information flow*, where systems do not explicitly describe the possible disclosures they can make; and *obscuring actual information flow*, where a system might not explicitly make clear what information is actually disclosed. The control pitfalls are: *emphasising configuration over action*, where configuration overshadows the privacy management actually needed to adapt to

a user's ordinary use of the system; *lacking coarse-grained control*, where a system offers too many choices and not just simple on/off choices; and *Inhibiting existing practice*, where a system forces some required practice onto a user, and does not adapt to the user's practice.

The ASTRA project does not try to minimise asymmetry in information flow. In fact, the underlying model explictly allows to model information flow where the originator of the information is not even aware that private information is being sent [6]. In contrast, privacy is achieved through introducing several methods of selectively making awareness states available. Work by Lederer et al. has shown that the "identity of the information inquirer is a stronger determinant of privacy preferences than is the situation in which the information is collected" without completely ignoring the influence of the situation [11]. This is supported by results from a study by Consolvo et al. who conclude that users wanted to "determine whether and what to disclose about their location to requests from social relations: *i.e., who* is requesting, *why* do they need to know, *what* would be most useful to them, and *am I willing* to share that?" [12]

Price et al. suggest a model for user control of privacy where they identify two main groups of methods for privacy protection, namely (1) *policy matching* which executes an explicit model of information exchange, and (2) *noise* where aspects of the user are disguised [13]. The authors continue to group noise into five different groups:

1. *Anonimizing:* hiding the identity of the user.
2. *Hashing:* disguising the identity of the user.
3. *Cloaking:* making the user invisible.
4. *Blurring:* decreasing the accuracy of the location (and possibly time).
5. *Lying:* giving intentionally false information about location or time.

For a social awareness system like ASTRA, the first two types are of limited use. This leads to three main aspects of privacy through *plausible denial* which have been identified by Metaxas and Markopoulos [6]:

Deception/Lying Intentionally supplying false information, for example by making contradictory information available to different communities.

Denial/Cloaking Hiding information, for example by concealing information about certain aspects (e.g. location).

Blurring/Evasion Only revealing part of the information, for example by decreasing the precision of the information provided, e.g. the location provided.

4 Explanations in Awareness Applications

Sørmo et al. [14] have earlier introduced a classification of explanations in intelligent systems. Most importantly, five different goals a user might have towards explanations are identified: The goal of *transparency* is concerned with the system's ability to explain how an answer was reached. *Justification* deals with the ability to explain why the answer is good. When dealing with the importance

of a question asked or action taken, *relevance* is the goal that must be satisfied. *Conceptualisation* is the goal that handles the meaning of concepts. Finally, *learning* is in itself a goal, as it teaches us about the domain in question. These goals are defined from the perspective of a human user.

We will in the following focus on the case that a user wants to get an explanation about his *focus* applications, e.g. we want to explore the relation between explanatory goals on one hand and privacy issue on the other. We have addressed the problem of explanations about the inner working of the system, for example by providing explanations about failure situations, already in earlier work [15]. On the nimbus part, that is when the user publishes information about himself, explanations will be of special interest when the user wants to determine the consequences of his actions, but we believe that the focus part highlights the most important relations.

Let us consider again the example given in the introduction, and how the system can satisfy the different user goals. We do not consider the learning goal, as this would only apply if we would use the ASTRA components to support learning about a different domain the users are involved in. We are focusing on the receiving part, e.g. the user who has subscribed to the awareness states of another user (Bob in our example above).

Justification: We expect the system to be able to justify why a certain action was taken. If we assume that Bob interacts with several different persons, he might have difficulties in remembering which colour displayed on his picture frame stands for which awareness state of a given user. So he can request to know why the frame shows a picture of Alice with a blue frame, and the system can give the justification that Alice's wish for a walk was made available and is to be displayed in this way.

Transparency: This goal shares similarities with the justification goal, but in contrast, it gives insights into the reasoning process. A transparency explanation for the display on Bob's picture frame would be given by displaying the particular rules leading to the event and by pointing out that the precondition – Alice's publication of her wish for a walk – was fulfilled. Another aspect would be that the system describes how it knows about Alice's wish for a walk.

Relevance: It is possible to define complex rules which take the recipients own context into account. For example, Bob can define that the picture frame should only be used when he is in his living room, but that a voice account should be given if he is in a different room of his flat. Therefore, if he sees that the picture frame has changed the colour, he might demand an explanation for why an audible signal was not given. The system would point out that the rule set specifically says that audible messages should only be given if Bob is not in the living room. To this end, it is not necessary to display the whole rule set, therefore the relevance explanation differs both from the transparency and justification explanation. We would like to point out that transparency explanations are especially useful for expert users who might want to modify the rule set, whereas the relevance and justification

explanations are especially useful for novice users or to give a quick overview about the actions taken by the system.

Conceptualisation: The system would deliver additional information about certain awareness states. In our example, Alice would have provided a textual description of what her wish for a walk means, this could be displayed to the user. In addition, the different awareness applications could be tagged with keywords, or described in terms of an ontology. It can be expected that different communities will develop different meanings for the same application, as conceptualisation is a distributed sense making process of all involved parties. Therefore, presenting the ontology for a given user community to Bob can help him better understand what certain awareness states mean. Likewise, presenting for example a tag cloud for a given community can help Bob understand which concepts are important in a given community. It can be expected that the tags used by a community of photo enthusiasts differ significantly from a community of co-workers, and a visualisation of the frequency of use can give a quick overview about the shared interests and conceptualisation of these different groups. For the time being, textual description of awareness applications is implemented in ASTRA, while tagging and ontology features are still under consideration.

Our previous research into the issue of designing explanation-aware systems was focusing on situations where the necessary knowledge is in principle available [15,16]. Such an omniscient system would defy the privacy protecting measures introduced in the ASTRA project. If the system would know that Alice was deliberately not making an aspect available, a transparency explanation to answer the question of why a certain action was not taken would reveal this fact to the user. In the ASTRA setting, we can therefore only generate explanations with the information we gain from looking at the general system status and the aspects made available to the individual user's focus. What implications does this have for explanations? Let us consider the different privacy protecting measures introduced in Section 3. We will focus solely on the *transparency* and *justification* goals since the deceptive techniques have no significant influence on *relevance* and *conceptualisation* explanations.

Deception/Lying: Let us consider that Alice had previously invited Carol to come over for a chat when she returns home, but she decides that she is too tired after having finished her paper. Since she had already cancelled meeting Carol a couple of times, she decides to tell her that she is busy at work, thereby intentionally supplying false information.

Carol has a focus application running telling her when Alice comes home by turning a specific lamp in her living room on. When this lamp is not lighted at the time she expects Alice to be home, she can ask for the following explanations:

Justification: The justification given will probably include the statement that the location aspect of Alice's nimbus is not set to home, in addition the system might tell where Alice says she is (if this information is in

general available to Carol). Although factually incorrect when looked at from the point of view of an all knowing observer, for Carol, this explanation will not be distinguishable from a true statement. Alice's privacy can only be endangered if other factual information contradicts this explanation.

Transparency: The same holds for transparency.

Denial/Cloaking: Let us now look at the possibility that instead of lying about her location, Alice decides to cloak information about her location from Carol. Let us further assume that she usually (or at least previously) made the information available to Carol. Asking for an explanation why the lamp is not lit will lead to the following situation:

Justification: The system will explain that the location aspect is not available on Alice's nimbus. This might endanger Alice's privacy since Carol is aware that such information was previously available (since she had her lamp coupled to this aspect). Although Alice's real location will not be available to Carol, she will have reason to belive that Alice is engaged in cloaking.

Transparency: The transparency explanation will lead to a similar result, but even more direct. While the justification explanation will deliver only the most important facts to support the explanation, leaving the conclusions to the user, a transparency explanation will point out the fact that the system could not determine Alice's location because that information was not available.

Blurring/Evasion: Dave is a co-worker of Alice and working with her on the paper. Alice is busy with preparing a meeting, but does not want to give Dave this information. Therefore, although she could make "preparing a meeting" available as her current activity, she decides to set her activity to working. Dave, who has a focus application running that informs him when Alice is working on the paper, might want to know why he does not get this information.

Justification: The justification information will give away the fact that Alice is working, but that no information about what she is working on is available. So Dave still gets information about Alice's activity, and this information will also not be contradicted by other available information (since preparing the meeting is a sub activity of working). Since writing the paper is also part of Alice's work, Dave will not gain any further information. Depending on the general social relation of the two, he might suspect that Alice is not working on the paper (if for example Alice usually gives very detailed information about her activity), but such an assumption is independent from the explanation delivered by the system.

Transparency: The same holds for transparency, with the slight distinction that the system probably will point out that writing the paper is a sub activity of working. Still, further reasoning will be left to Dave.

5 Discussion and Future Work

In this paper, we have looked at how explanation goals can be applied to social awareness systems. We have outlined how explanations can in general support the user of such systems, and we have sketched which explanation goals the ASTRA project can support on the focus part. We have also discussed how privacy issues affect the explanations that can be given by the system. The usefulness of such explanations might be limited compared to what we have suggested earlier [15], but this is due to (1) the necessity to support the privacy mechanism implemented in ASTRA and (2) the limited information available to our explanatory mechanism. In general, we can say that the privacy issues outweigh the explanation needs.

For future work, it would be interesting to perform a more thourough comparison of an omniscient explanation mechanism, e.g. one that can inspect the nimbus of other users with the capabilities of the asynchronous model on which ASTRA privacy is based. It would furthermore be interesting to compare that to explanation engines which can inspect the nimbus of other people, but adhere to the principle of minimising asymmetry.

Another aspect which has to be addressed in the ASTRA project in general, but which is of special importance for the issue of explanation, refers to work around ontologies and folksonomies. It would be interesting to see whether a bottom up approach via tagging can lead to sufficient explanations, or whether a strict ontology oriented approach is necessary.

The last aspect is not only valid for social awareness system, but for ambient intelligent systems in general. What is the semiotics of behavioural interfaces, and how can they integrate explanatory capabilities?

Acknowledgements

Part of the research reported in this paper is financed by the EU project ASTRA-EU029266.

References

1. Astra: Project website. Internet (2008) Last visited 2008-04-18.
2. Marsh, S.: Exploring the socially adept agent. In: Proceedings of the First International Workshop on Decentralized Intelligent Multi Agent Systems (DIMAS 1995). (1995) 301–308
3. Pieters, W.: Free will and intelligent machines. Project Report, NTNU Trondheim (2001)
4. Benford, S., Bullock, A., Cook, N., Harvey, P., Ingram, R., Lee, O.K.: From rooms to cyberspace: models of interaction in large virtual computer spaces. Interacting with Computers **5** (1993) 217–237
5. Rodden, T.: Populating the application: A model of awareness for cooperative applications. In: CSCW '96: Proceedings of the 1996 ACM conference on Computer Supported Cooperative Work, New York, NY, USA, ACM Press (1996) 87–96

6. Metaxas, G., Markopoulos, P.: 'aware of what?' a formal model of awareness systems that extends the focus-nimbus model. In: Proceedings of the IFIP conference EHCI 2007, Springer (2007)
7. Langheinrich, M.: Privacy by design – principles of privacy-aware ubiquitous systems. In Abowd, G.D., Brumitt, B., Shafer, S.A., eds.: Proceedings of the Third International Conference on Ubiquitous Computing (UbiComp 2001). Number 2201 in Lecture Notes in Computer Science, Springer (2001) 273–291
8. Bellotti, V., Sellen, A.: Design for privacy in ubiquitous environments. In Michelis, G.D., Simone, C., Schmidt, K., eds.: Proceeding of the Third European Conference on Computer-Supported Cooperative Work (ECSCW '93), Kluwer Academic Publishers (1993) 77–92
9. Jiang, X., Hong, J.I., Landay, J.A.: Approximate information flows: Socially-based modeling of privacy in ubiquitous computing. In Borriello, G., Holmquist, L.E., eds.: Proceedings of the 4th International Conference on Ubiquitous Computing (UbiComp 2002). Volume 2498 of Lecture Notes in Computer Science., Springer (2002) 176–193
10. Lederer, S., Hong, I., Dey, K., Landay, A.: Personal privacy through understanding and action: five pitfalls for designers. Personal and Ubiquitous Computing **8** (2004) 440–454
11. Lederer, S., Mankoff, J., Dey, A.K.: Who wants to know what when? privacy preference determinants in ubiquitous computing. In: CHI 2003 extended abstract on Human factors in computing sysems. (2003) 724–725
12. Consolvo, S., Smith, I.E., ad Anthony LaMarca, T.M., Tabert, J., Powledge, P.: Location disclosure to social relations: why, when & what people want to share. In: Procedings of the SIGCHI conference on Human factores in computing systems. (2005) 81–90
13. Price, B.A., Adam, K., Nuseibeh, B.: Keeping ubiquitous computing to yourself: A practical model for user control of privacy. International Journal of Human-Computer Studies **63** (2005) 228–253
14. Sørmo, F., Cassens, J., Aamodt, A.: Explanation in case-based reasoning – perspectives and goals. Artificial Intelligence Review **24** (2005) 109–143
15. Kofod-Petersen, A., Cassens, J.: Explanations and context in ambient intelligent systems. In Kokinov, B., Richardson, D.C., Roth-Berghofer, T.R., Vieu, L., eds.: Modeling and Using Context – CONTEXT 2007. Volume 4635 of Lecture Notes in Computer Science., Roskilde, Denmark, Springer (2007) 303–316
16. Cassens, J., Kofod-Petersen, A.: Designing explanation aware systems: The quest for explanation patterns. In Roth-Berghofer, T.R., Schulz, S., Leake, D., eds.: Explanation-Aware Computing – Papers from the 2007 AAAI Workshop. Number WS-07-06 in Technical Report, Vancouver, BC, AAAI Press (2007) 20–27