CPS^2: A Contextual Privacy Framework for Social Software

Rula Sayaf[1](✉), Dave Clarke[1,2], and Richard Harper[3]

[1] Department of Computer Sciences, IMinds-DistriNet, KU Leuven, Leuven, Belgium
rula.sayaf@cs.kuleuven.be
[2] Department of Computer Sciences, Uppsala University, Uppsala, Sweden
dave.clarke@it.uu.se
[3] Microsoft Research, Cambridge, UK
r.harper@microsoft.com

Abstract. Social software has become one of the most prominent means for communication. Context is essential for managing privacy and guiding communication. In social software, context can be ambiguous due to the overload of data and the mix of various audiences. Such ambiguity may result in privacy issues.

To overcome context and privacy issues, we propose CPS^2, a conceptual framework for contextual privacy management. The frameworks is based on an analysis of the role of context in communication and privacy management. The analysis identifies the interpretation of data as a key ingredient for privacy management. We present CPS^2 and how the preservation of interpretation within any context facilitates preserving contextual privacy. We discuss how CPS^2 can be technically realised, and how it can address context issues and offers fine-granular context control.

Keywords: Context · Privacy · Social software · Data interpretation · Communication · Contextual privacy

1 Introduction

Communication through social software is becoming one of the most prominent ways of daily communication. Social software is an application for the exchange of various types of data to communicate with a large number of users. Such communication is simple as it can be achieved by disclosing data to other users. This simplicity can be associated with privacy issues. Issues specifically occur when data is accessed by an inappropriate audience or put in inappropriate contexts [1]. To mitigate such issues, users should have means to control context to ensure appropriateness and preserve their privacy.

Context is essential for both communication and privacy management [2,3]. Context is the information that characterises situations. Context facilitates the interpretation of data [4]. In communication, context facilitates clarifying and

J. Tian et al. (Eds.): SecureComm 2014, Part II, LNICST 153, pp. 25–32, 2015.
DOI: 10.1007/978-3-319-23802-9_4

delivering the communicated message [2]. Through communication, the interlocutors express their identity through the data they disclose. Managing one's identity is the core aspect of privacy as informational self-determination [3,5]. By controlling context, privacy can be managed to manage one's identity. When context is unclear and ambiguous, communication can be disrupted affecting one's identity expression and privacy. Context ambiguity can be seen in social software. Ambiguity is caused by the mixing of different audiences and data from different contexts. As a result, privacy and communication can be affected.

Managing privacy through controlling context is a complex task. Controlling context requires reasoning about the current context and how it may change [6]. Such reasoning is challenging due to the high-dimensionality of context parameters [4]. Current context-based privacy management approaches address such complexity by simplifying context representation resulting in a limited control over context [7]. To understand the insufficiency of context-based management consider the following scenario that is based on a reported incident of 'prostitutes of Antwerpen' [8]:

Scenario 1. Els is a fashion model, and she posts her photo in a swimming suit on Facebook and makes it public. Els experiences a privacy issue when her photo is disseminated in the context of 'prostitutes of Antwerpen' page, which affects her job applications. In contrast, Els does not face any issue when her photo is disseminated in 'jobs for top models' context.

Current privacy management approaches do not offer sufficient context control to mitigate the violations mentioned in the scenario. Most approaches do not offer the possibility to allow appropriate disseminations and prohibit the inappropriate ones. They can only either allow all disseminations or prohibit them. In this paper, we address such context control issues by proposing a conceptual framework for contextual privacy management. We analyse the context-privacy relation and argue that the interpretation of data is a key ingredient in contextual privacy management. We demonstrate that by ensuring the integrity of interpretation, contextual privacy can be managed. The framework is a conceptual approach to manage privacy in context without burdening users with reasoning about context and its complexities. The contributions of this paper are the following:

1. Analysing the problems of controlling data and managing privacy in a context-based manner (Sect. 2)
2. Analysing the role of context in privacy management and communication (Sect. 3)
3. Proposing a conceptual framework for Contextual Privacy for Social Software (CPS^2), and presenting how this framework can be technically realised and can address context and privacy issues (Sect. 4).

2 Problem Statement

Communicating while preserving privacy in any context requires a fine-grained control of context [7]. In social software, context identifies situations where

various types of data are disclosed and users interact. Context ambiguity is one of the main issues in social software communication. Ambiguity means that it is challenging to accurately identify the current context. Ambiguity obstructs the clarification of the communicative message, and user's assessment of privacy.

Privacy management can be challenging due to context management problems. Privacy is viewed as the means to control the contexts in which data is put [9,10]. According to this view, contextual privacy management requires two types of control: control over the *original* context in which the data was originally disclosed through the software, and control over dissemination contexts by specifying appropriate or inappropriate contexts in which data can be put or not, respectively. Practicing these two types of control is complicated. A user can control the original context by choosing where to disclose data and to whom. However, over time, the original context might change [6] into an inappropriate context. In order to avoid such situations, users should constantly monitor changes. Often, users do not invest much time in managing and monitoring online communication contexts [11], and it is particularly challenging when context is ambiguous. Having control over any dissemination context requires listing possible appropriate or/and inappropriate contexts, depending on the assumed closed- or open-world of contexts. Given the 'theoretically infinite complexity' of social situations, and the infinite set of possible contexts [12,13], it may be infeasible to list all possible contexts [14]. Context issues are often insufficiently addressed by simplistic context representations in privacy management approaches. Such approaches offer limited context control [7] that is insufficient to satisfy privacy and communication requirements in social software.

3 Analysis of Context and Privacy

Context is the information construct that characterises the communication situation [4]. Context is a container of data; it facilitates the inference of the relevant meaning of the communicative message [4]. A data item can have a set of different possible meanings or interpretations, and by identifying the context it is put within, the relevant interpretation can be inferred. An example is the context of the page in which Els's photo is put in Scenario 1. That this context is related to 'prostitutes' can be inferred by the information about the type of page, content, creator, and other meta data. When Els's photo is put in this context, the most relevant meaning of the photo is a 'prostitute_photo'.

In online communication, privacy management can be a means of identity management [5]. The data owner[1] discloses a data item to communicate about it with the selected audience. Through communication, the owner expresses a specific identity and manages it by specifying who the audience are and what data they could access in a specific context [15]. To make the privacy decision of to whom disclose an item, the owner estimates how others would perceive and interpret this item [5]. Thus, the interpretation of data and context are of central roles in the process of privacy management.

[1] We do not imply the legal ownership.

The importance of context and the interpretation of data can be mainly observed in two communication types: cooperative and adversarial. These types are the extreme ends of the communication spectrum, and are characterised by variant degrees of trust, context involvement, and privacy concerns [16]. In *cooperative communication*, the interlocutors trust each other [17] and act jointly to understand and interpret the communicated message. Cooperative communication can be achieved by following the Gricean maxims, which concern providing a sufficient amount of information that is true, relevant, and unambiguous to make context explicit [18]. Gricean maxims facilitate clarifying the context to make possible interpreting the communicated message. However, in ambiguous contexts, it is challenging to abide by those maxims. In contrast, in an *adversarial communication*, at least one of the interlocutors—the adversary—can violate Grecian maxims to mislead others into misinterpreting the message and disrupt the communication. Adversarial communication is associated with low trust and high privacy concerns [16]. In both communication types, context ambiguity hinders the correct interpretation of data affecting the identity expression and privacy of the interlocutors.

Based on the above-mentioned argument, we define contextual privacy management as the process of managing data disclosure or dissemination while maintaining the appropriate interpretation of this data, in order to manage the one's desired identity in a specific context. To achieve that, context clarity is essential. However, clarity of context requires an effort to make communication cooperative and avoid adversarial communication. To facilitate contextual privacy management and avoid overloading users with context complexities, we propose CPS^2 in the following section.

4 CPS^2: Contextual Privacy for Social Software

The main idea of CPS^2 is to facilitate communication with an increased level of privacy without burdening users with context management. We propose CPS^2 to manage contextual privacy by maintaining the appropriateness of the interpretation of data. CPS^2 avoids simplifying the representation of context or imposing reasoning about context on users to specify privacy management policies. Given the technological advances in context inference [19] and automatic data interpretation [20], CPS^2 does not requires users to reason about context, rather, it requires owners to only specify the appropriate interpretation of their data. The framework is responsible for guarding the appropriate interpretation upon any change of context or dissemination, as explained in the following.

To understand the principle of CPS^2 consider Scenario 1: Els's profession as a fashion model is indicated on her page, thus, the context of her profile page indicates that the 'fashion-related' interpretation is the most relevant interpretation. Upon viewing the photo, the audience would highly likely perceive the interpretation of the photo as such. When the photo is put in the 'prostitutes' context, the relevance of the 'fashion-related' interpretation is low and the relevance of the 'prostitute' interpretation is high, which affects Els's identity.

With CPS^2, Els can specify the set of appropriate interpretations of the photo as {fashion_show, swim_suits_show, pretty_model}. Accordingly, the dissemination into the 'prostitutes' context should be prohibited because it results in an interpretation that is not in the set Els has specified, while the dissemination into the 'jobs for top models' context should be allowed.

4.1 Realisation of CPS^2

The realisation of CPS^2 implies a system with three main functions: context inference, interpretation inference, and contextual privacy management. CPS^2 assumes the existence of an underlying context inference and interpretation inference layers that need not be managed by users, but by the social software provider, for instance. The realisation would comprise the following layers:

1. Context inference layer: responsible for inferring or labelling the context of the current situation within the social software realm. The input to this layer is the social software data: users and their attributes, data items, relations, ads, and the structure of its pages and modules. When data is added to a situation, this layer adapts and infers the new context.
2. Interpretation inference layer: responsible for inferring the interpretation of data based on the context inferred by the previous layer. The data can be interpreted whether it is textual or visual.
3. CPS^2 control layer: responsible for facilitating contextual privacy management by means of two possible approaches: access control or accountability and auditing approach. The access control approach comprises a policy language to express the contextual privacy policies and an enforcement mechanism. A policy can be formulated to express the appropriate interpretations of a data item. Upon performing an action—resulting in adding or removing data from a context—the control layer consults the policies of data items in the current context and verifies the appropriateness of the interpretation inferred by the previous layer. The action is executed if no interpretation is inappropriate.

   In the accountability and auditing approach, users need not specify policies. Rather, upon a context change, the framework marks the actions that cause a change of the interpretation. The data owner can verify the appropriateness of a new interpretation. If the new interpretation is inappropriate, proper actions can be executed against the responsible entity.

4.2 Addressing Issues of Context and Privacy

CPS^2 could potentially address the problems mentioned in Sect. 2, as follows:

1. Context ambiguity: the framework addresses this problem not by making the context less ambiguous to users, rather, even when context is ambiguous to some users, the context inference layer could still identify context given all the data in the software. Accordingly, only appropriate actions are allowed.

2. Context simplistic representation: by shifting the burden of reasoning about context to the underlying framework, it is not needed to simplify the representation of context.
3. Control over the original context: by facilitating the management of interpretation, owners can indirectly control context to a relatively high degree without having to monitor the changes of context.
4. Control over any context: the previous argument is valid here. The framework facilitates effortless control over any context by continuously monitoring and maintaining the appropriateness of the interpretation in any context.

Moreover, CPS^2 enhances communication to become cooperative even if context is ambiguous, by allowing only appropriate actions that may not affect the interpretation of data. It also facilitates avoiding adversarial communication by preserving data interpretation. CPS^2 facilitates control over data flow in both private or public spaces.

5 Related Work

Many works have incorporated context in privacy management. On the conceptual level, Nissenbaum proposes contextual integrity [21] for privacy management. She presents a list of norms: contexts, actors, attributes, and transmission principles, that must be managed to preserve privacy. Our framework differs from this theory by not requiring an exhaustive specification of the possible contexts or the other ingredients of the theory. The complexity of contextual integrity results in models that adopt simplistic context representation to overcome the complexity. An example is the formal model of Barth *et al.* [22] where context is represented by roles of users.

Another contextual privacy management work is Fong's access control model. In his work, relationships are viewed as contexts [23]. In contrast to CPS^2, Fong's model offers control over the original context but not over dissemination contexts. Generally, the simplification of such models reduces the granularity offered by context and fails in addressing the problems discussed in Sect. 2.

6 Conclusion and Future Work

In CPS^2, we propose maintaining data interpretation to manage contextual privacy and address the complexity of controlling context. The framework facilitates simple management of privacy without reducing the richness context management offers. CPS^2 enhances communication in which interpretation is essential. In other work, we have conducted experiments related to context inference, and we will report them elsewhere. Our future work aims at providing a design to validate the framework and investigate the proper realisation of the framework.

Acknowledgment. This research has been funded by the IWT in the context of the SBO project on Security and Privacy for Online Social Networks (SPION). Thanks are due to Natasa Milic-Frayling and Sören Preibusch at Microsoft Research Cambridge.

References

1. Goldie, J.: Virtual communities and the social dimension of privacy. University of Ottawa Law & Technology Journal **3**(1), 133–167 (2003)
2. Clark, H., Carlson, T.: Context for comprehension. In: Long, J., Baddeley, A. (eds.) pp. 313–330. Lawrence Erlbaum Associates, Inc, Hillsdale (1981)
3. Gürses, S.: Multilateral privacy requirements analysis in online social network services. Ph.D. thesis (2010)
4. Van Dijk, T.A.: Discourse and context. A Sociocognitive Approach. Cambridge University, Cambridge (2008)
5. Palen, L., Dourish, P.: Unpacking "privacy" for a networked world. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems pp. 129–136. ACM (2003)
6. Mcculloh, I.: Detecting changes in a dynamic social network. Ph.D. thesis, Carnegie Mellon University (2009)
7. Sayaf, R., Clarke, D.: Access control models for online social networks. Social Network Engineering for Secure Web Data and Services, 32–65 (2012)
8. De Wolf, R.: Over 'spotted', 'hoeren' en 'failed'-pagina's. Electronic article (2013). http://www.knack.be/nieuws/belgie/dader-antwerpse-hoeren-foto-geklist/article-4000230766578.htm, Last checked February 2013
9. Westin, A.: Privacy and Freedom. Atheneum, New York (1970)
10. Petronio, S.: Boundaries of privacy: Dialectics of disclosure. SUNY Press, Albany (2002)
11. Lipford, H.R., Besmer, A., Watson, J.: Understanding privacy settings in facebook with an audience view. In: Proceedings of the 1st Conference on Usability, Psychology, and Security, Berkeley, CA, USA, pp. 2:1–2:8. USENIX Association (2008)
12. Van Dijk, T.A.: Context models in discourse processing. In: The Construction of Mental Representations During Reading, pp. 123–148 (1999)
13. Skantze, G.: Error Handling in Spoken Dialogue Systems-Managing Uncertainty, Grounding and Miscommunication. Doctoral dissertation, KTH. Ph.D. thesis, Department of Speech, Music and Hearing (2007)
14. Lampinen, A., Lehtinen, V., Lehmuskallio, A., Tamminen, S.: We're in it together: interpersonal management of disclosure in social network services. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3217–3226, ACM (2011)
15. Wood, A.F., Smith, M.J.: Online Communication: Linking Technology, Identity, & Culture. Routledge, London (2004)
16. Harper, R.H.: Texture: Human Expression in the Age of Communications Overload. MIT Press, Cambridge (2010)
17. Harper, R. (ed.): Trust, Computing and Society. CUP, New York (2014)
18. Grice, H.P.: Logic and conversation. In: Davidson, D., Harman, G. (eds.) The Logic of Grammar, pp. 64–75. Harvard Univ., Cambridge (1975)
19. Cao, H., Hu, D.H., Shen, D., Jiang, D., Sun, J.T., Chen, E., Yang, Q.: Context-aware query classification. In: Proceedings of the 32nd international ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 3–10. ACM (2009)
20. Celikyilmaz, A., Hakkani-Tur, D., Tur, G.: Statistical semantic interpretation modeling for spoken language understanding with enriched semantic features. In: Spoken Language Technology Workshop (SLT), 2012 IEEE, pp. 216–221. IEEE (2012)

21. Nissenbaum, H.: Privacy in context: technology, policy, and the integrity of social life. Stanford Law & Politics, Stanford (2010)
22. Barth, A., Datta, A., Mitchell, J.C., Nissenbaum, H.: Privacy and contextual integrity: framework and applications. In: IEEE S & P'6, pp. 184–198. IEEE Computer Society (2006)
23. Fong, P.W.L.: Relationship-based access control: protection model and policy language. In: Proceedings of the first ACM Conference on Data and Application Security and Privacy. CODASPY 2011, New York, NY, USA, pp. 191–202. ACM (2011)