Saveface: Save George's faces in Social Networks where Contexts Collapse

By Fuming Shih and Sharon Paradesi

fuming@csail.mit.edu, paradesi@csail.mit.edu

Massachusetts Institute of Technology

1. Introduction

Social networks have repeatedly perplexed Web architects with the dilemma of privacy versus publicity. Is it ever possible to protect privacy on a platform that exists and is designed to share user's information publicly? Despite the efforts of increasing privacy controls for its users, Facebook has continuously made the headline news concerning privacy issues [7]. Seemingly, more controls to a user's content does not afford a satisfactory solution to diminish privacy risks [8]. Studies have shown that it is unreasonable for a user to foretell all the possible consequences of information disclosure and set an ex ante rule about privacy [10]. The real challenges about privacy protection in social networks, as some researchers argue, are the dynamics of social contexts [4], and the deficiency of current architecture in utilizing contextual information [2].

Through this paper, we demonstrate a need to create a sustainable architecture to create and implement viable privacy awareness measures and practices for different actors within the social network 'eco-system'. Also, by monitoring transfer of information and social actions within the system, the system would be able to suggest the policy decisions being made to the user.

The rest of the paper is structured as follows. We explain the meaning of context within social networks, the limitations in the current social networks and finally highlight research directions that will help in implementing a sound privacy infrastructure. Our paper primarily focuses on Facebook and Flickr, but we envision that these principles are universal in nature.

2. Contexts in Social Networks

In Social Networks, contexts imply a content's original intent, the original audience addressed, content's object references, the original activity or practice; and context tied to social or public in which the content is produced [5]. In general, when a user reads a piece of information, she usually places it within an implicit context in order to interpret it. Multiple contexts together act as a "container" of the disclosed information to differentiate it from other similar kinds. Without defining contexts, the consumers of the information have no reference to appropriately evaluate the message conveyed.

In the famous Seinfeld episode *The Pool Guy* [12], George Costanza was upset when confronted with

the situation of his new girl friend infiltrating his group of friends. In the past, he had successfully kept his different relationships (friends, love and so on) in independent silos, and did not expect them to mix together. Most Facebook users today face a similar dilemma. When their moms and bosses "friend" their way to access a user's content, privacy concerns arise due to the lack of context supported by the Facebook architecture.

Another dimension to the issue of privacy is the dichotomy of data present in social networks - facts versus gossip. Facts can be described as content that explicitly references the user(s) mentioned in them, while gossip does not. A privacy-honoring system should treat privacy as a function of user's expectations in particular attributable contexts [1].

3. Current limitation - lack of context in a social networks

Neglect of contexts when processing information could cause systems to disregard user's expectations and thereby raise serious privacy issues [13]. In the real-world practice of privacy management, companies have gradually shifted their focus from content to the context within which the user's data is present. This change aims to make appropriate use of users' data to meet their expectation [1]. To prevent users from experiencing George's panic, the social network should incorporate user's contexts explicitly so that application that plug into it would treat information per user's specifications. On the other hand, only a user herself could best evaluate the privacy risk based on the effects brought by the change to her context.

For example, George could specify a policy that states that all comments attached to pictures from or tagged by certain friends (for instance, those he parties with) should not be used for employment decisions or dating recommendations. Once George states this context within which his data should be interpreted, smart Facebook applications that implement "employee-finder" systems or other external programs that mine Facebook data [14] should bypass certain pieces of information when aggregating data about George that they would otherwise utilize. By setting such restrictions on contexts, George will still be able to receive personalized services like the targeting ads but is now able to state when certain information cannot be used against him.

Such a mechanism is currently lacking in Facebook and we think that the directions described in the next two sections would lead to the development of infrastructure to satisfy this lack.

4. Vision towards a more efficient privacy setting in social networks

a. Creating an awareness about contexts in user

A formal description of "contextual integrity" was developed by privacy law researcher Helen Nissenbaum who introduced an inextricable relation between privacy and context [2]. However in the realm of a social network, it is unreasonable to ask the user to manually identify the relevant contexts for their privacy concerns. To alleviate the user's burden, the author in [6] proposed an privacy mechanism to learn a user's privacy policy based on context inference. We envision that it is important to help a user explore relevant contexts to raise the awareness of possible privacy risks

Say that George updated his friend list to include Susan. We can give him a warning about the content that Susan can see before and after becoming his friend. In other words, we show George how his worlds will collide. George is now in a better position to decide whether to create a new group for Susan or not.

b. Need for more flexible access controls

Most of the current social networks provide some level of access control. For example, Facebook and Flickr provide default settings of control where a user can restrict certain people from viewing her content. However, in reality, having well-defined groups of friends is a very naive assumption from the perspective of the dynamism of human relationships. A more reasonable way that we propose is to enable users to create dynamic groups based on attributes that matter to the user. These attributes would then form the context within which the friends, or even strangers, would be placed.

For instance, say that George uploads content on Facebook hoping to reach a large audience to convey some message. This audience could be larger than any group he currently has but he does not want to manually create a new list of friends just for this update. In such a scenario, a flexible way would be to obtain the attributes that George wants his audience to have and automatically create such a list for him.

c. Augmenting access control with usage restrictions

Relying solely on access control can only get you so far. In extreme cases, access control only makes users paranoid about who is viewing and using their information. Currently, one can only restrict access to his or her content, but has no say over how that content will be used. We would like to see the current infrastructure progress to an ideal world where everyone can exercise their freedom of speech. To reach such a goal, we envision that users should be able to attach usage restrictions of how they want the data to be interpreted and used by others.

In short, if Elaine uploads content about George on her Facebook account and tags George, all his usage restrictions should be applicable on this content as they would carry over to data that is beyond his ownership.

d. Standardizing usage restrictions across multiple networks

Implementing usage restrictions in a single network seems straightforward. However, the real challenge comes when we need to reconcile these restrictions across multiple networks and media [11]. The crucial aspect is that the restrictions should be tied to and travel along with the piece of information.

For example, say that George uploads a photo of himself on his Flickr account and Elaine posts a note about it on her Facebook account. If George was tagged in the said note by Elaine, his friends can view it. In such a situation, George's usage restrictions from Flickr should carry over to the note pasted by Elaine on Facebook as the provenance of the photo is Flickr.

5. Project Saveface: current status and future work

The goal of Saveface is to represent and store the social network world within the context of a user. To achieve this, we created ontologies to describe user data created on the Social Web (as of now, Facebook

and Flickr). With these ontologies, we crawl all the content owned by the user and the user's friends on Facebook and Flickr as viewable by the user when logged in. The crawled content is then saved locally as RDF data. Each piece of information in the RDF graph can be visualized as a node in a graph. Different nodes in the graph are linked by relationships that reflect the way data is connected in Facebook.

For example, George has a friend named Elaine and both signed up for an event mentioned on a Facebook Page. Now, both George and Elaine appear as nodes on the graph which are related with "friend relationship" as described by Facebook. The event they both signed up for appears as another node on the graph that is linked to their respective nodes with a link titled "attending".

One advantage with this data model is that a user can explore her social graph without the current limitations presented in the user interface of Facebook or Flickr. A user could query the social graph using SPARQL to fetch data within a specific context. For example, George could ask for all the comments that Elaine posted on his wall about photos tagged with "party".

We propose the following technical implementations to achieve the four-fold vision mentioned above. To tackle,
a. Context awareness, we plan to develop an interactive interface for people to retrospectively view privacy settings that are in conflict with their expectations.
b. Flexible access controls, we propose to develop a data miner that understands the attributes that the user is looking for and match relevant users to those attributes. A sample attribute could be "those working in the same lab where I worked during the last year".
c. Usage restrictions, we propose an interface that allows a user to set usage restrictions on her data similar to that mentioned in [9]. One way to do that would be to create a "named graph" [3] that defines the usage restrictions of an object to be honored by those accessing it.
d. Usage restrictions across multiple networks, we would like to implement data structures that allow attaching usage restrictions to the data that travels across social networks as "sticky" policies.

6. Conclusion

As we have seen in this paper, most access control mechanisms are not sufficient for a dynamic and evolving environment like a social network. To improve, we proposed a four-fold vision to attack this problem and discussed the technical details of how to get there.

References

[1] K. A. Bamberger and D. K. Mulligan, Privacy on the books and on the ground. Stanford Law Review 63 (2010). Accessed 14 May 2010.
[2] A. Barth, A. Datta, J. Mitchell, and H. Nissenbaum, Privacy and Contextual Integrity: Framework and Applications, Proceedings of the IEEE Symposium on Security and Privacy, 2006.
[3] C. Bizer, J.J. Carroll, P. Hayes, P. Stickler.Named Graphs, Provenance and Trust. Proceedings of the 14th international conference on World Wide Web, 2005.
[4] D. Boyd 2008. Facebook's Privacy Trainwreck: Exposure, Invasion, and Social Convergence. Convergence, 14 (1), 2008.
[5] A. Chan, Social context, Facebook Likes, activity and action streams, entry posted April 27, 2010,

http://www.gravity7.com/blog/media/2010/04/social-context-facebook-likes-activity.html
[6] G. Danezis. Inferring privacy policies for social networking services. In AISec, 2009.
[7] M. Helft, Facebook Acknowledges Privacy Issue With Applications, entry posted October 18, 2010,
[8] L. Kagal and H. Abelson. Access control is an inadequate framework for privacy protection. In W3C Privacy Workshop, 2010.
[9] T. Kang and L. Kagal. Enabling Privacy-awareness in Social Nework. In Intelligent Information Privacy Management, AAAI Spring Symposium Series, 2010.
[10] A.M. McDonald, R.W. Reeder, P.G. Kelley, and L.F. Cranor. A comparative study of online privacy policies and formats. Privacy Enhancing Techonologies Symposium 2009.
[11] Z. Wu, L. Wang: Enforcement of Privacy Policies over Multiple Online Social Networks for Collaborative Activities.SCSS 2009: 583-588
http://bits.blogs.nytimes.com/2010/10/18/facebook-admits-to-privacy-issue-and-makes-fixes/?hp
[12] Independent George - Worlds collide, http://www.youtube.com/watch?v=uPG3YMcSvzo
[13] R. Wauters, The Washington Post, Google Buzz Privacy Issues Have Real Life Implications . http://www.washingtonpost.com/wp-dyn/content/article/2010/02/12/AR2010021201490.html
[14] M. Elegan, Datamation, 'Pre-crime' Comes to the HR Dept. http://www.socialintelligencehr.com/home

Acknowledgements

The authors would like to thank Hal Abelson (from MIT), Joe Pato (from HP labs and currently at MIT) and Michael Speciner (from MIT) for their insightful comments and discussion.