Modeling Social Networking Privacy

Carolina Dania*

IMDEA Software Institute, Madrid, Spain
Universidad Complutense de Madrid, Spain
carolina.dania@imdea.org

Abstract. Privacy-related issues are becoming a serious concern among users of social networks. There are at least three reasons that justify this growing concern: social networking privacy policies are hardly trivial; they live in a constant state of flux; and they are only informally and partially described by the social networking sites.
To improve this current state of affairs, we propose SecureUML as a formal language to model social networking privacy, and we set ourselves the goal of modeling, as a case study, Facebook privacy policy. Based on our formal model, we envision a new generation of tools that will provide Facebook users with more information about the privacy of their posts and about the associated privacy-related risks.

1 Motivation

Many people in our society rightly consider themselves as "internet natives": when they need information, they naturally open a browser and search for it; when they want to share information, they naturally post it on a social network. A few figures about Facebook, the leader among social networking sites, exemplify our point: Facebook has more than 800 million users, of which, about 50% log on to their accounts every day; more to the point, Facebook users upload, on average, 250 million photos per day [5].

Not surprisingly, privacy-related issues are a growing concern among users of social networking sites [7, 1, 14, 15] and, consequently, among their developers. Last November, Facebook's founder and CEO, Mark Zuckerberg, wrote in his blog [16] "I also understand that many people are just naturally skeptical of what it means for hundreds of millions of people to share so much personal information online, especially using any one service. Even if our record on privacy were perfect, I think many people would still rightfully question how their information was protected." Then, recognizing an increasing criticism over Facebook privacy policy, Zuckerberg announced: "we're making a clear and formal long-term commitment to do the things we've always tried to do and planned to keeping doing —giving you tools to control who can see your information and then making sure only those people you intend can see it."

To Facebook's credit, over the past 2 years, its users have been equipped with new tools and resources which are designed to give them more control over

* This work has been partially supported by the EU-NoE project NESSoS, 256980.

their Facebook experience, including: an easier way to select your audience when making a new post; inline privacy control on all your existing posts; the ability to review tags made by others before they appear on your profile; a tool to view your profile as someone else would see it; many more privacy education resources.

Despite all these efforts, many users are still concerned about how to maintain their privacy or —in Zuckerberg's words— "rightfully questions how their information was protected". There are at least three reasons for this:

- The Facebook privacy policy is hardly trivial to understand: for example, when tagging policies and privacy settings conflict to each other.
- The Facebook privacy policy has been in a constant state of flux over the past few years [12], and it is prompted to change again in the near future.
- The Facebook privacy policy is only informally and partially described in a collection of "privacy education resources", which cannot provide a coherent and complete account of the policy.

As a consequence, even advanced Facebook users may find difficult to understand, for example, the actual visibility of a post. To illustrate our point, recall the tagging policy explained in [6]:

> "*When I tag someone in a photo or post, who can see it?* When you tag someone, it may be visible to: 1. The audience you selected for your post. 2. Friends of the person you tagged (if the audience is set to "Friends" or more). (...) *When someone adds a tag to a photo or post I shared, who can see it?* When someone adds a tag to something you shared, it's visible to: 1. The audience you chose for the post or photo. 2. The person tagged in the post, and their friends (if the audience is set to "Friends" or more)."

Now, suppose that Bob, Alice, Ted, and Peter have Facebook profiles: Bob is friend of Alice and Ted; Ted is friend of Peter; Ted is not a friend of Alice; and Peter is not friend of Alice or Bob. Assume also that Alice has set to "Friends" the default audience for posts of friends in her wall. Consider the following scenarios:

Scenario #1 Alice posts a photo of herself, Bob and Ted in her wall, and set its audience to "Friends". Then, Bob tags Ted in this photo. *Question: Can Peter see this photo in Alice's wall?* (The answer is Yes.).

Scenario #2 Bob posts a photo of himself, Ted and Alice in Alice's wall. Then, Bob tags Ted in this photo *Question: Can Peter see this photo in Alice's wall?* (The answer is No. Why?).

2 Research Project

Objectives. We set ourselves two goals: first, to provide a formal account of the Facebook privacy policy (as complete as possible); and second, to design methods (based on this formal account) for reasoning about sharing and privacy in Facebook.

Potential impact. We envision at least three new Facebook privacy tools that can use our results as a solid, rigorous foundation: first, a tool for checking whether a person can see a post (currently, this tool is only available for the owner of the wall where the post is posted, but not for the creator of the post); second, a tool for assessing the risk of a post becoming visible for a person; and third, a tool for assessing the impact, on the visibility of a post, of a default privacy policy change. We also expect our methodology to be applicable to other social networking site, like Google+, opening the path for a formal comparison between privacy policies of different social networking sites.

Context. This project is being conducted at IMDEA Software (`http://software.imdea.org`) Modeling Lab, under the co-supervision of Manuel Clavel and Marina Egea. Manuel Clavel is Associate Researcher at IMDEA Software and Professor at Universidad Complutense de Madrid. Dr. Marina Egea is Consultant & Research Project Manager at Atos S.A., Spain. We are developing this research within the European Network of Excellence for Engineering Secure Software and Systems.

Methodology. As discussed in [13], when modeling social networking privacy it is crucial to use a language able to formalize *fine-grained* access control policies: in other words, a role-based access control language, as proposed in [9], will only do part of the job. There are differents options for this, but not so many when having a formal semantics becomes a hard requirement. For example, XACML [10], which can be considered the standard choice for describing privacy policies, lacks of a formal semantics.

To provide a formal account of the Facebook privacy policy, we use SecureUML [3]. SecureUML is a formal language for modeling role-based access control. It provides a rich language for expressing both static and dynamic access control policies, the latter being policies that depend on the run-time satisfaction of authorization constraints. Based on our preliminary results, we believe that SecureUML is up for the task we have set to ourselves for the following reasons:

1. Facebook ultimately decides about the visibility of a post based on the settings chosen by the owner of the wall and on the relationships (if any) that link the visitor of the wall, the owner of the wall, the creator of the post, and the creators and targets of the tags (if any) added to the post. Interestingly, when only real users are considered (i.e., no Facebook-enhanced games, applications, websites, or advertisers) the *purpose* of the visitor (and, similarly, for the creator of the post or of the tags) play no role in Facebook decisions; neither assigns the Facebook privacy policy any *obligation* to the visitor.
2. Facebook's profiles, walls, posts, photos, tags, and so on, can be naturally modeled in SecureUML as *entities*, with the expected relationships between them: the owner of a wall, the creator of a post, the wall where a post is posted, the post where a tag is added, and so on. In particular, privacy settings can be modeled as *attributes* of the entities 'profile' and 'post' while the relationship of friendship can be modeled as a *self-association* of the entity 'profile'.

3. Facebook's policy constraints (like 'a user can read a post if he or she is a friend of the owner of the wall where the post is posted') are naturally modeled in SecureUML using OCL [11]. OCL is a strongly typed, declarative language, specifically designed for querying scenarios consisting of entities (with attributes) and associations between them. In particular, using OCL, we can (the list is by no means exhaustive):
   - refer to the value, in a data element, of any of the attributes specified in the data model.
   - refer to all the data elements which are linked to a data element through any of the associations specified in the data model.
   - perform standard operations on booleans (negation, conjunction, disjunction, implication, etc.).
   - perform equalities/inequalites between (collection of) data elements of the same type.
   - perform standard operations on collections (union, intersection, emptiness, inclusion, exclusion, insertion, deletion, etc.).

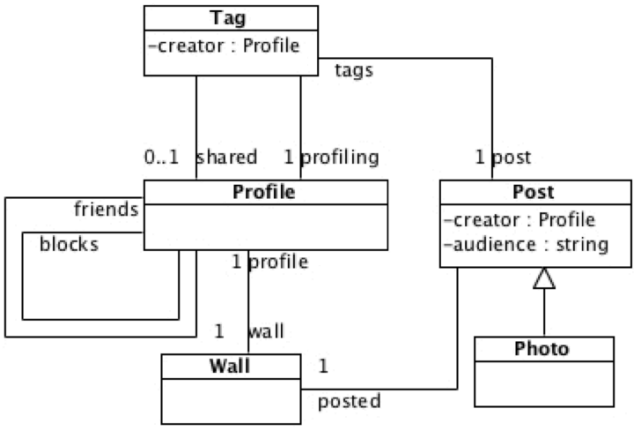

Fig. 1. Data model for Facebook posts and tags (partial).

To illustrate our methodology, we show in Figure 1 a basic data model that (partially) represents Facebook posts and tags. Using the entities, attributes, and associations contained in this data model, we show in Figure 2 how the following clauses —about when a visitor (@caller) can read a post (@post)— are formalized in OCL:

- anybody can read any post that is posted in his or her wall, independently of its creator.
- anybody can read any post that is posted in a wall when he or she is a friend of the owner of the wall and the audience selected is 'Friends'.

```
@caller=@post.posted.profile
or (@post.audience='Friends' and @post.posted.profile.friends->includes(@caller))
or (@post.posted.profile.blocks->excludes(@caller) and
        (@post.audience='Public'
         or @post.tags.profiling->includes(@caller)
         or (@post.audience= 'Friends' and @post.creator=@post.posted.profile
              and @post.tags.profiling.friends->includes(@caller))
)) or ...
```

Fig. 2. Authorization constraints for reading a Facebook post (partial).

- anybody can ready any post that is posted in a wall when the audience selected is 'Public', unless he or she is blocked by the owner of the wall.
- anybody can read any post that is posted in a wall when he or she is tagged in this post, independently of the audience selected, unless he or she is blocked by the owner of the wall.
- anybody can read any post that is posted in a wall, when the audience selected is 'Friends', he or she is a friend of somebody tagged on the post, he or she is not blocked by the owner of the wall, and the owner of the post happens to be the creator of the post.

On the other hand, with respect to our second goal (namely, reasoning), SecureUML has a well-defined semantics that supports the formal analysis of its models. In particular, for a certain type of analysis, we can automatically analyze SecureUML models using the metamodel-based approach described in [2]. For more general analysis, we can use theorem-proving tools (including SMT solvers), thanks to the mapping from OCL to first-order logic introduced in [4].

3 Research Plan

Our first task is to gather as much information as possible about the Facebook privacy policy. We would like to assume that the information available at [6] is correct and complete. Unfortunately, our initial results show that this is not always the case. Our second task is then to design "experiments" on precooked Facebook scenarios for testing that our understanding of the Facebook privacy policy corresponds to reality. Eventually, these "experiments" should also help us to monitor and report changes in the Facebook privacy policy. We will also look very closely at the results of the thorough and detailed audit [8] of Facebook's practices and policies by the Office of the Irish Data Protection Commissioner.

According to the gathered information, we will decide how to proceed. We envision separated tasks for modeling each of the basic actions on Facebook: select audience, switch reviews on/off, read a post, write a post, remove a post, add a tag, remove a tag, approve a tag, add a friend, remove a friend, block a user, and so on. For each of these actions, we plan to model also their pre- and post-conditions using OCL. We are aware that we may have to exclude from our

project the privacy policies that apply to advertisers and/or so-called Facebook-enhanced games, applications, and websites, unless we obtain this information directly from the company. Finally, we plan to design methods (based on the different types of analysis for SecureUML models that we mentioned before) for reasoning about sharing and privacy in Facebook (e.g., audience evaluation, risk and change impact assessment), although the actual implementation of these methods may have to be carried out in other research projects.

References

1. A. Acquisti and R. Gross. Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In G. Danezis and P. Golle, editors, *Privacy Enhancing Technologies*, volume 4258 of *LNCS*, pages 36–58. Springer, 2006.
2. D. Basin, M. Clavel, J. Doser, and M. Egea. Automated analysis of security-design models. *Information and Software Technology*, 51(5):815–831, 2009.
3. D. Basin, J. Doser, and T. Lodderstedt. Model driven security: From UML models to access control infrastructures. *ACM TOSEM*, 15(1):39–91, 2006.
4. M. Clavel, M. Egea, and M. A. G. de Dios. Checking unsatisfiability for OCL constraints. *Electronic Communications of the EASST*, 24, 2009.
5. Facebook. `http://www.facebook.com/press/info.php?statistics`.
6. Facebook. Facebook Help Center. `http://www.facebook.com/help`.
7. R. Gross, A. Acquisti, and H. J. Heinz. Information Rvelation and Privacy in Online Social Networks. In V. Atluri, S. D. C. di Vimercati, and R. Dingledine, editors, *WPES*, pages 71–80. ACM, 2005.
8. Irish Data Protection Commissioner. Facebook Ireland Ltd. Report of Audit, December 2011.
9. J. Li, Y. Tang, C. Mao, H. Lai, and J. Zhu. Role based access control for social network sites. In *Pervasive Computing (JCPC), 2009 Joint Conferences on*, pages 389 –394, dec. 2009.
10. OASIS. eXtensible Access Control Markup Language (XACML), 2010. `http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-cs-01-en.pdf`.
11. Object Management Group. *Object Constraint Language specification Version 2.2*, Feb. 2010. OMG document available at `http://www.omg.org/spec/OCL/2.2`.
12. N. O'Neill. Infographic: The History of Facebooks Default Privacy Settings. `http://www.allfacebook.com/`.
13. A. Simpson. On the Need for User-Defined Fine-Grained Access Control Policies for Social Networking Applications. In *Proceedings of the workshop on SOSOC*, pages 1:1–1:8, New York, NY, USA, 2008. ACM.
14. A. Young and A. Quan-Haase. Information Revelation and Internet Privacy Concerns on Social Network Sites: A Case Study of Facebook. In *Proceedings of the international conference on C&T*, pages 265–274, New York, NY, USA, 2009. ACM.
15. E. Zheleva and L. Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In J. Quemada, G. León, Y. S. Maarek, and W. Nejdl, editors, *WWW*, pages 531–540. ACM, 2009.
16. M. Zuckerberg. Our Commitment to the Facebook Community. The Facebook Blog, Nov. 2011. `https://blog.facebook.com`.