Privacy in Evolving Social Networks

Raúl Pardo[1] and Gerardo Schneider[2]

[1] Dept. of Computer Science and Engineering, Chalmers, Sweden.
[2] Dept. of Computer Science and Engineering, University of Gothenburg, Sweden
{pardo@chalmers.se, gerardo@cse.gu.se}

1 Motivation

Over the past decade, the use of the social networks like Facebook and Twitter, just to mention two of the most popular ones, has increased at the point of becoming ubiquitous. Many people access *Social Networks Services* (SNSs) on a daily basis; e.g. to read the news, share pictures with their friends or check upcoming events. Nearly 70% of the Internet users are active on SNSs, as shown by a recent survey [4]. Empirical studies have also shown that the current privacy protections offered by SNSs are very far from the users' expectations [6, 5]. One of their weaknesses is the inability for users to express desirable privacy policies. This is because, the privacy settings offered by SNSs are too coarse-grained. Furthermore, many users are not fully aware of the result of activating a privacy policy or if the policy protects their personal data as they expect. We believe citizens should be in power to control and decide on how much information to make public. One way to do so is by providing users with means to define their own privacy policies and give guarantees that they will be respected. Privacy in SNSs may be compromised in different ways: from direct observation of what is posted (seen by non-allowed agents), by inferring properties of data (*metadata privacy leakages*), indirectly from the topology of the SN (e.g., knowing who our friends are), to more elaborate intentional attackers such as *sniffers* or *harvesters* [3]. Among others, one of the origins of these attacks comes from their privacy enforcement mechanism, the so called Relationship-Based Access Control (ReBAC) [2].

We aim at developing a privacy enforcement mechanism which offers social network users the possibility of expressing finer-grained privacy policies, enabling them to deal with (certain kinds of) implicit disclosure of sensitive information. Moreover this mechanism should take into account that the social network *evolves* and enforce the privacy policies as events occur. We have developed the privacy policy framework $\mathcal{FPPF}$ for social networks [8, 7], which is briefly described in next section.

2 The First-Order Privacy Policy Framework $\mathcal{FPPF}$

$\mathcal{FPPF}$ is composed by a static part which describes the state of the social network at a given point in time, and the dynamic part, which models how the social network evolves as events occur [7]. The components are:

A social network model. SNM is a *social graph*, a graph whose nodes represent users, and edges represent different kind of relationships between users. The graph is enriched with information on the knowledge the users of the social network have, and what they are permitted to do.

A knowledge-based logic. $\mathcal{KBL}_{\mathcal{SN}}$ is an epistemic logic including a permission operator, which provides the possibility to reason about what the agents know and what they are allowed to do. The logic allows us not only to access and reason about the explicit knowledge of an agent, but also about implicit knowledge (through inferences).

A formal privacy policy language. $\mathcal{PPL}_{\mathcal{SN}}$ is a language for writing privacy policies for each individual user.

A labelled transition system for social networks. $\mathcal{LTS}_{\mathcal{SN}}$ contains a set of SNMs, which result from the execution of events in the SNSs. These events as operational semantics rules, which are divided in the following categories: i) *Epistemic.* These rules describe how the knowledge and the permission change in the SNM. ii) *Social topology.* These rules modify the social topology of the SNM, i.e. the users and their relationships. For instance, adding new users, relationships between them, etc. iii) *Policy.* These rules allow for the modification of the set of privacy policies of the agents. iv) *Hybrid.* These are rules which combine changes of any of the categories above.

Besides, the framework also comes with a *satisfaction relation* defined for the logic $\mathcal{KBL}_{\mathcal{SN}}$, and a *conformance relation* defined for the policy language $\mathcal{PPL}_{\mathcal{SN}}$. The framework may be tailored by providing suitable instantiations of the different relationships, the events, the atomic predicates representing what is to be known, and the additional facts or rules a particular social network should satisfy.

In order to show how $\mathcal{FPPF}$ can be used, we have instantiated the privacy policies of Facebook and Twitter [8, 7], which are two of the most used social networks nowadays. For instance, one of Facebook's privacy policies is responsible for setting the audience of a post, where the user can choose among 'Friends', 'Only me' and 'Custom'. In $\mathcal{FPPF}_{\text{Facebook}}$ it would be split in 3 policies. In the mentioned instantiation if u wants the audience of her posts to be her Friends, it would be written as follows: $[\![\neg S_{Ag\setminus friends(u)\setminus\{u\}} post(u,j,n)]\!]_u$ where $S_G\varphi$ is a formula stating "somebody in the group G knows φ", Ag is the set of all the agents in the SNM, $u, j \in Ag$, $n \in \mathbb{N}$, $post(u,j,n)$ represents post n, written by u and posted in j's timeline and $friends(u)$ is an function which returns all the friends of u.

As we mentioned before $\mathcal{FPPF}$ is an generic framework, therefore we could combine instantiations of two (or more) different social networks in one. This is a very useful and innovative feature, since currently it is becoming more common to connect several accounts from different social networks and share information between them. As a final example of the use of our framework, we present below an example of a privacy policy concerning the combination of $\mathcal{FPPF}_{\text{Twitter}}$ and $\mathcal{FPPF}_{\text{Facebook}}$ [8]. The following privacy policy: *Only my friends in Facebook who are following me in Twitter can know my location* will be written in our formalism as $[\![\neg S_{Ag\setminus(friends(u)\cap Followers(u))\setminus\{u\}}\ u.location]\!]_u$. The rules defining how the SNM evolves are given using small step operational semantics. For example, in Twitter the most basic event is called *tweet*. It is used to share a 140 characters long message with a set of users. Let $tweet(tu, TweetInfo, Audience)$ be the event representing that user tu shares $TweetInfo$ to the set of users $Audience$. The following rule models the behaviour of the event,

$$\frac{\forall \varphi \in TweetInfo, \forall i \in Audience\ KB_i' = KB_i \cup \{\varphi\}}{SN \xrightarrow{tweet(tu,TweetInfo)} SN'}$$

We use $SN \xrightarrow{tweet(tu,TweetInfo)} SN'$ to denote that an SNM SN evolves to a new SNM SN'. $TweetInfo$ is a set of formulae in $\mathcal{KBL}_{\mathcal{SN}}$, which represents the content of the message, e.g. if the predicate $age(u) \in TweetInfo$, it means that u's age is part of the message. KB_i represents the *knowledge base* of a user i. In SNMs knowledge bases are used to store all the information that the users know. Let KB_i' be the knowledge base of i in SN' and analogously for KB, then $\forall \varphi \in TweetInfo, \forall i \in Audience\ KB_i' = KB_i \cup \{\varphi\}$ means that after the execution of the event *tweet*, all users part of the audience will *know* all the information shared in the tweet.

In [7] we define what means for a SNS to be privacy-preserving. Specifically, we say if all privacy policies are executed before and after the execution of any event in the SNS, then the

SNS is privacy-preserving. We also proved that Twitter is privacy-preserving. Additionally, we proved that adding new desirable policies to Twitter and Facebook make the SNSs not privacy-preserving.

3 Current and Future Work

Traditionally the semantics of "the logic of knowledge" or epistemic logic are given by means of Kripke structures, where the *uncertainty* of the agents in modelled [1]. However, in $\mathcal{FPPF}$ we explicitly model the concrete knowledge of the agents. While these approaches to model knowledge seem to be complementary, we recently found out that it is possible to encode SNM to the equivalent Kripke structure. Nevertheless, this encoding entails some issues that we are currently investigating. For instance, the properties of the binary relations in the Kripke structure affect the implicit knowledge that agents can infer. It might be possible that some properties of knowledge that hold in Kripke structures do not hold in SNMs. Besides, we plan to investigate whether the models are equivalent, i.e. we claimed that it is possible to encode SNMs in Kripke structures, but we do not know if the opposite is possible.

Besides, we are currently implementing a prototype of our framework in the open source SNS *Diaspora**[1][2]. We have extended Diaspora* with several privacy settings, which are not currently offered in other major SNS (including Facebook or Twitter). We aim at implementing the full power of $\mathcal{FPPF}$. A centralised implementation of the enforcement mechanism would create a huge bottleneck, since SNSs are massively distributed and millions of events could be triggered at the same time. Therefore, we are currently looking into distributed architectures for monitoring algorithms, which can help us to monitor the privacy policies of all users in the SNS efficiently.

As future work we plan to extend $\mathcal{FPPF}$ to support real time policies. For example, a user could write a policy saying "My boss cannot know my location between 20:00-23:00" or "The audience of the post on my timeline during my birthday is only my friends".

References

[1] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge*, volume 4. MIT press Cambridge, 2003.

[2] P. W. Fong. Relationship-based access control: Protection model and policy language. In *CODASPY'11*, pages 191–202. ACM, 2011.

[3] B. Greschbach, G. Kreitz, and S. Buchegger. The devil is in the metadata - new privacy challenges in decentralised online social networks. In *PerCom Workshops*, pages 333–339. IEEE, 2012.

[4] A. Lenhart, K. Purcell, A. Smith, and K. Zickuhr. Social media & mobile internet use among teens and young adults. *Pew Internet & American Life Project*, 2010.

[5] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: User expectations vs. reality. IMC '11, pages 61–70. ACM, 2011.

[6] M. Madejski, M. Johnson, and S. Bellovin. A study of privacy settings errors in an online social network. (PERCOM Workshops'12), pages 340–345.

[7] R. Pardo, M. Balliu, and G. Schneider. A formal approach to preserving privacy in social networks (extended version). Technical report, Chalmers University of Technology.

[8] R. Pardo and G. Schneider. A formal privacy policy framework for social networks. In *SEFM'14*, volume 8702 of *LNCS*, pages 378–392. Springer, 2014.

[1] `https://diasporafoundation.org/`

[2] `https://github.com/raulpardo/ppf-diaspora`