Trust-aware Privacy Control for Social Media

Na Li
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
na.li@epfl.ch

Maryam Najafian Razavi
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
maryam.najafian-razavi@epfl.ch

Denis Gillet
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
denis.gillet@epfl.ch

Abstract
Due to the huge exposure of personal information in social media, a challenge now is to design effective privacy mechanisms that protect against unauthorized access to social data. In this paper, a trust model for social media is first presented. Based on the trust model, a trust-aware privacy control protocol is proposed, that exploits the underlying inter-entity trust information. The objective is to design a fine-grained privacy scheme that ensures a user's online information is disclosed only to sufficiently trustworthy parties.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Keywords
Privacy, trust, social media, Web 2.0

ACM Classification Keywords
H5.0. Information interfaces and presentation: General.

General Terms
Design, Human Factors.

Introduction
Along with the success of Web 2.0 social media, privacy control over the shared information has been receiving growing attention in recent years. Most Web 2.0 platforms use the simple notion of "friends" or "connections" for privacy control, which is too coarse to allow any fine-grained data management. As a result, so far there is no simple way to grant access rights to only a set of trusted parties, to disclose a specific collection of artifacts to selected audiences, or to specify what actions are allowed on certain content. Furthermore, social media enables people who do not

necessarily know each other to interact and exchange information. Trust-based privacy policies allow users to specify their privacy preferences based on their trust relationships and as such, can improve users' experiences with social media by helping them to regulate and control their social data while interacting with others on the Web.

In this paper, we propose a trust-aware privacy control approach that allows users to control the exposure of their personal and collaborative data based on their trust relationships. The objective is to take advantage of the available trust information to make sure that data is disclosed only to sufficiently trustworthy parties. In the rest of the paper, three key dimensions of privacy control in social media are addressed first. Then, a trust model for social media is presented and a trust-based privacy protocol that accounts for all the three aforementioned dimensions is introduced. Finally, the major benefits and potential problems of the proposed protocol are discussed, along with a discussion of future directions.

The framework

Users of social media produce a wide variety of user-generated content, such as profile information, blogs, comments, and photos, which together create one's life-long online identity. Anwar et. al define privacy as a user's capacity to control the conditions under which her identity information will be shared [1]. One may wish to disclose different partial identity information to different audiences. For instance, one might be willing to share her party photos with her close friends, while only exposing limited personal information to the friends of her friends. An effective privacy control scheme should enable users to specify which piece of information they are willing to expose to whom in which way. To this end, three key dimensions are taken into account in the proposed privacy control approach: **audience**, **action**, and **artifact**.

The **audience** of information could be a person, an application, or anything else that can access user-generated content in social media (e.g., widgets, services, and so on). Previous studies have shown that the audience of information plays an important role in users' sharing behavior [5]. As such, a main aspect of every privacy management mechanism must be enabling users to define specific audience (including both people and/or applications) for their various online information.

In addition to defining audience, users should also be able to specify what **actions** the authorized parties can perform on the disclosed data: one may grant editing permission over a collaborative document to her co-workers, allow other colleagues to only view the content, and keep the document inaccessible to strangers.

The **artifact** dimension represents any type of data element that is created by a user and could be shared in social media, including user profile attributes, posted resources, comments, messages, and so on. Evidently, it is necessary for users to have control over the specific data that would be exposed to others. One might keep her sensitive personal information like test scores confidential, while making her movie interests publicly visible.

The approach

Based on these requirements for privacy management, we now present a trust-aware privacy control approach that tackles all three dimensions outlined above.

Using trust to inform privacy

Most privacy frameworks share similar dimensions to define the privacy problem (i.e., there is always the question of how to specify *who* should be allowed to perform *what* action on *which* artifact). However, different approaches for categorizing the audience dimension have been proposed in the literature. In OpnTag system [5] for example, people-tags allow the artifact owner to define the audience for her data based on her relationship with the receiver (e.g., colleague), or her assessment of him/her (e.g., expert). Others have proposed audience categorization based on other factors, including degree of closeness to the information requester [2], and type of relationship between the sender and receiver of information (i.e., personal vs. professional) [4].

The purpose of this paper is to investigate whether the inter-entity trust relationship can be used as a viable basis for audience categorization. The motivation behind the idea is that intuitively, people feel more comfortable to share personal information with trustworthy parties than unknown ones. A recent study also identified trust as a main factor in users' information sharing decisions [5]. Inspired by these observations, we propose to categorize different audiences for shared information based on the trust relationship between the owner and the receiver of information. To quantify inter-entity trust relationship in social media, a trust model is introduced first.

A trust model for social media

Users of social media naturally express their trust opinions in both **explicit** and **implicit** ways. By adding a person as a friend or blocking someone, one **explicitly** indicates her trust or distrust opinions with regard to that user. Similarly, adding a particular application into the trusting list or blacklist shows the explicit trust level a user has in that application.

Moreover, actions performed by users of social media result in heterogeneous types of relationships like tagging, linking, membership, commenting, or rating [3]. Those relationships **implicitly** suggest different amounts of potential trustworthiness depending on the importance of that particular type of relationship. For instance, the action of Alice positively rating Bob's post indicates that Alice has a certain degree of trust in Bob. We proposed in [3] a multi-relational trust metric that aims at measuring the implicit trust relationship between a target user and other parties in her trust network. The basic idea of the proposed trust metric can be briefly described as follows.

The trust value that is derived from a particular type of relationship is defined as **Direct Trust**. Let R_i denote a relationship of type i existing between a user s and another party t, $W(R_i)$ denote the weight of relationship R_i, and $N(s,i)$ denote the number of outgoing relationship edges of type i from the user s. Then $DT(s,t)$, the Direct Trust value of the user s regarding the party t, can be inferred as in (1):

$$DT(s,t) = \frac{W(R_i)}{N(s,i)} \quad (1)$$

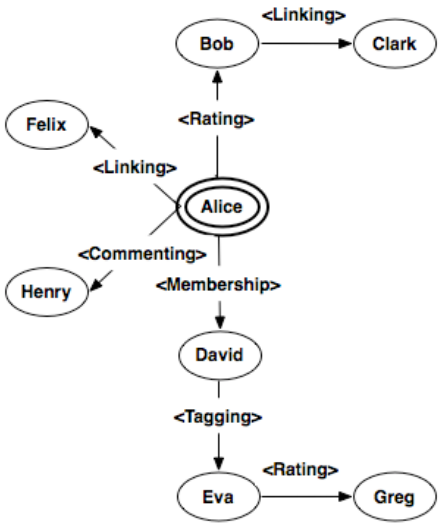

figure 1. Alice's Trust Network.

Besides the Direct Trust, trust could also propagate along the relationship path starting from the target user, as it does in real life. For instance, if Alice trusts Bob and Bob trusts Clark, Alice may have a certain amount of trust in Clark. The trust value that is derived from trust propagation through relationship path is defined as **Indirect Trust**. However, trust relationship is not completely transitive, and it could decay through distance. Therefore, a **Propagation Distance** of trust is introduced to constrain the range that trust is able to propagate (i.e., trust relationship is unable to extend beyond that distance). Based on the Direct Trust derived from the social relationships and Indirect Trust derived from trust propagation, a trust network of a particular user is constructed within the trust propagation distance.

Figure 1 illustrates the trust network of a user, called Alice. Alice joined David's ski club, David tagged Eva in his photo, and Eva rated Greg's article. The relationship path indicates an implicit trust propagation from Alice to David, then to Eva, and finally to Greg. Similarly, trust also propagates through other social relationships starting from Alice, and a personalized trust network of Alice is generated accordingly.

Using the Direct Trust value between each pair of entities, Indirect Trust value can be inferred by extending the trust network layer by layer, centered on the target user. The trust values of the user's direct neighbors are computed first, followed by computing the entities at the second distance level. The trust inference process is continuously performed until it reaches the predefined trust propagation distance. The inferred trust value for an entity at a certain distance is the average of all the incoming trust edge values, weighted by the trust value of the corresponding entity that the trust edge is derived from. Let s denote the target user that lies at the center of the trust network, and t denote an entity at a certain distance in s's trust network. E represents the set of all the entities e_j that has a direct trust edge to t. $T(e_j, t)$ denotes the trust value from e_j to t, and $T(s, e_j)$ denotes the trust value from s to e_j. Then the Indirect Trust value from s to t, $IT(s,t)$, is inferred as in (2):

$$IT(s,t) = \frac{\sum_{e_j \in E} T(e_j,t) T(s,e_j)}{\sum_{e_j \in E} T(s,e_j)} \qquad (2)$$

For a particular user, the implicit trust values of all the entities in her trust network can be inferred using the multi-relational trust metric described above. As mentioned previously, a user might have indicated explicitly her trust opinions in some users or applications. To deal with the conflict between explicit and implicit trust values of a particular entity, we adopt **Explicit-Applicable** policy that gives higher priority to users' explicit trust opinions. For instance, if a user adds an application into her blacklist, the application is considered as totally untrusted even if its implicit trust value is high.

Trust-aware privacy control

In order to provide a fine-grained privacy scheme that allows users to specify their privacy preferences based on their trust relationships, we introduce the notion of **Privacy Protocol** that takes into account all the three aforementioned dimensions necessary to control privacy for information sharing.

A privacy protocol, defined by the owner of a piece of information, is a set of rules each declaring a categorization of audiences that is permitted to perform a set of actions over a set of artifacts. The person or the application requesting access to a particular artifact should be verified to satisfy the privacy protocol before actually accessing that artifact. Each rule in the privacy protocol consists of three elements: audience control element, action control element, and artifact control element, corresponding to the three key privacy dimensions respectively. Each element is discussed in detail hereafter.

Based on the observation that people tend to share more information with trustworthy parties than unknown ones, we propose to define the audience dimension based on the existing inter-entity trust values derived from our proposed trust model. To this end, the audience control element in the privacy protocol can be represented using a **Trust Barrier** value, referred to as TB. The trust barrier specifies the lower bound on the trust value that a group of audience must have in order to be authorized to access a particular artifact. In other words, only people or applications with a trust value higher than or equal to the trust barrier are granted access. The trust value of a person or an application can be inferred using the trust metric proposed in the previous section. From the perspective of the owner of a piece of information, data is disclosed only to sufficiently trustworthy parties.

As far as the action control element is concerned, we define a notion of **Action Set**, referred to as AS, which is a set of access actions such as viewing, editing, deleting, linking, tagging, rating, and commenting. The action set constrains the access actions that the audience is allowed to perform over a particular artifact.

With regard to the artifact control element, users are able to assign a **Confidentiality Level** (referred to as CL) to all their artifacts. For instance, one might keep her salary information as the "most confidential" artifact, while specifying her general profile information as "not confidential". This facilitates organizing one's artifacts according to different protection purposes.

Finally, a privacy protocol (referred to as PP) consisting of the three elements discussed above could be represented as in (3). An illustrative example could be PP = (0.9, {Editing, Commenting}, "Medium Confidential"), which means "People with trust values higher than or equal to 0.9 can edit and comment on my artifacts with confidentiality level of 'Medium Confidential'". This privacy protocol efficiently prevents the untrusted people (with trust values less than 0.9) from accessing (editing and commenting) a given collection of artifacts (with medium confidentiality level).

$$PP = (TB, AS, CL) \tag{3}$$

Discussion

The proposed privacy control approach provides a trust-based solution for managing the over exposure of personal information in social media. One of the major benefits of this approach is that it is fine-grained. It allows specifying a categorization of trusted audiences that could perform a particular set of actions over a collection of artifacts. Furthermore, the approach enables users to define their own privacy preferences and organize their online information according to

different privacy requirements. Users have full control over the disclosure of their social data. Finally, since the inter-entity trust information and the three key dimensions of privacy management already exist in most of the social media platforms, the proposed privacy solution is applicable to a variety of social systems.

Although the trust-aware privacy control approach can potentially be an effective solution to social data protection, it also poses several challenges. Due to the complexity of privacy protocols, designing a usable implementation that enhances user experience with privacy management could be challenging. It is not clear how big the usability issues are if a user is faced with the request to define such fine-grained privacy policy statements. Besides, as the approach relies on the available inter-entity trust information derived from users' online activities, providing adequate incentive schemes that motivate users' participation should also be considered.

Conclusion and future work

To protect users' online information, we have proposed, a trust-aware privacy control approach that takes advantage of the underlying inter-entity trust information in social media to inform privacy. Trust relationship between entities is quantified in both explicit and implicit ways. Based on that, a privacy protocol consisting of three key dimensions (audience control, action control, and artifact control) is constructed to ensure that data is only disclosed to sufficiently trustworthy parties.

We are currently in the process of implementing the proposed privacy approach in a social media prototype called Graaasp (graaasp.epfl.ch). The usability and acceptability of the privacy solution will be evaluated through user studies. Moreover, we believe that the trust-based policies could be used not only for privacy management, but also for filtering social noise. An interesting direction could be to apply trust-based restrictions over one's received information to make sure that only content from trusted senders would be received.

Reference

[1] Anwar, M.M., Greer, J., and Brooks, C.A. Privacy Enhanced Personalization in E-learning. In *Proc. PST 2006*, ACM Press (2006), 1-4.

[2] Consolvo, S., Smith, I. E., Matthews, T., Lamarca, A., Tabert, J., and Powledge, P. Location Disclosure to Social Relations: Why, When, & What People Want to Share. In *Proc. CHI 2005*, ACM Press (2005), 81-90.

[3] Li, N., El Helou, S., and Gillet, D. Trust-based Rating Prediction for Recommendation in Web 2.0 Collaborative Learning Social Software. In *Proc. ITHET 2010*, IEEE Press (2010), 197-201.

[4] Patil, S. and Lai, J. Who Gets to Know What When: Configuring Privacy Permissions in An Awareness Application. In *Proc. SIGCHI 2005*, ACM Press (2005), 2-7.

[5] Razavi, M. N. and Iverson, L. Designing for Privacy in Personal Learning Spaces. *New Review of Hypermedia and Multimedia, Special Issue on Studying the Users of Digital Education Technologies: Theories, Methods, and Analytical Approaches 13*, 2 (2007), 163-185.