Addressing the privacy management crisis in online social networks

Krishna Gummadi, *MPI-SWS* Balachander Krishnamurthy, *AT&T Labs–Research*
Alan Mislove, *Northeastern University*

Online social networks (OSNs) have become the defacto portal for Web access for millions of users, resulting in fundamental shift in the patterns of context exchange over the Web. Previously, content on the Web was primarily created by a (relatively) small group of publishers, including companies, universities, and governments. This content was (for the most part) public, with open sharing and universal access to information being explicit goals. Today, much of the content shared on the Web is being created by individual end users, and the increasingly personal nature of this information is causing privacy and access control to become key concerns.

The result of this fundamental shift is that instead of just being content *consumers*, individual end users are now required to be content *managers*. Today, for every single piece of data shared on sites like Facebook—every wall post, photo, status update, friend request, and video—the uploader must decide which of his friends, group members, and other Facebook users should be able to access the data. Given the per-user average of 130 friends and 80 groups—compounded with the average 90 pieces of content uploaded per user per month—it is unsurprising that we are in the midst of a *privacy management crisis*, wherein the task of simply managing access to their content has become a significant mental burden for many users.

In this position paper, we argue that this situation is being exacerbated by the lack of meaningful and intuitive privacy controls and abstractions. Instead of making the task of privacy management easier, the most common controls today require users to expend significant mental effort. The result is that most users simply do not use the controls at all: Facebook recently revealed that only 5% of users had *ever* created a friend list, one of the core mechanisms for expressing privacy. While others have taken this lack of use to indicate that users are no longer concerned about privacy, we believe that it is more symptomatic of the inadequacy of the controls themselves.

Problems with the state-of-the-art

In this section, we outline a number of problems with the state-of-the-art in privacy management.

Lack of proper access control mechanisms The most common access control mechanisms (e.g., friends, and friends-of-friends, friend lists) are primitive and often insufficient to capture user intent. For example, the set of friends-of-friends (the only grouping that includes non-friends but not the entire world) can range from hundreds to tens of thousands of users, and sites today leave users to guess at its true size. Moreover, mechanisms like friend lists impose significant burden on users, who are tasked with dividing up their friends and then required to maintain these lists over time.

In large part, the lack of proper access control mechanisms is because the mechanisms are inspired by the bygone era where computation was largely in the domain of corporations with well-defined hierarchical groups of users. Data sharing in social networks works in different ways, with overlapping and ever-changing social relationships.

Misunderstanding implications Users today often do not understand the implications of their actions and access control settings. For example, Facebook default privacy settings are so complex (and err on

the side of open access) that many users are often not aware of who can see their data. To make matters worse, different social networking applications that get access to users' data can expose it to others in ways users might not expect. The underlying problem is that users are expected to keep both the (ever-changing) privacy model and their privacy settings in their mind at all times, and to use this to make continual access control decisions.

The way forward

Unfortunately, we do not believe that there is any silver bullet to addressing privacy management problems in OSNs. The desires of individual users are often varied and conflicting, and are dependent on social relationships that even sociologists and psychologist have yet to understand. However, we believe that there are a few basic technological approaches that would represent first steps towards addressing the privacy management crisis.

Convey implications of actions OSNs need to convey to users the implications of their actions. For example, when then post a status update or upload a photo, how many other people are able to access that piece of information? Today, users must manually keep track of the privacy model and their privacy settings, using these two items to infer this set of users. Instead, providing the set of users is an operation that OSNs could trivially provide and would relieve the users of significant mental burden.

Provide privacy mirrors OSNs need to make it easier for people to understand who can see their data. To do so, we propose that OSNs provide *privacy mirrors*, which shows a user how his profile and data appears to other users in the network. In addition, OSNs need to provide users with a list of all of the different views that exist for their data.

Automatically infer groups We believe that the basic mechanism provided by friend lists (enabling the sharing of data to a subset of a user's friends) is a useful one. However, the implementation today places a significant burden on the user. As an alternative, we propose to help the user by automatically inferring "groups" from the structure of the social network (and, of course, allowing the user to manually edit this inferred list). We believe this to be a promising approach, as the structure of the social network has been shown to correlate strongly with the social groups users form. The advantage of automatically creating these lists is that it would relive the user of (a) setting up the initial lists, and (b) maintaining correct list membership over time.

As a first step, we have created a prototype implementation of this mechanism, deployed it to individual users, and have seen promising initial results. From an individual user's perspective, the local groups (approximating many desired friend lists) are clearly visible, and can be enumerated with off-the-shelf community detection algorithms. Going forward, the challenge is to capture the missing members of the lists and to extend the detection to two-hop friends and beyond. Clearly, more research needs to be done here.

Move from access control to exposure Traditionally, researchers and developers tend to think in terms of access control (i.e., for a given piece of data, who has access to that data). But, in OSNs, it may help to think in terms of *exposure*. For example, even though many users many have access to a piece of data, most may not view it, meaning the information is not widely exposed. On the other hand, status updates are automatically broadcast to a user's friends, greatly increasing exposure.

In moving from access control to exposure, we see that the problem can be significantly simplified. Instead of requiring the user to state, up front, the entire set of other users who are able to access a given piece of content, we could simply ask the user "How widely exposed should this piece of content be?" Once the piece of content has been viewed by that number of users, it is then hidden. This provides an understandable guarantee to users without requiring significant mental burden.

Infer appropriate exposure level It may even be possible to infer the expectation a user has about his or her data exposure from looking at the past history of accesses to the user's data or those of similar users. For example, if there is an avalanche of requests for a user's data on a single day, it might be worth warning the user. Similarly, if there are a significant number of requests that show unusual access patterns (e.g., to old data posted by the user several years ago or from others who are far away in the network), then it may be worth warning the user as well.