Generating Default Privacy Policies for Online Social Networks

Eran Toch
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213 USA
eran@cs.cmu.edu

Norman M. Sadeh
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213 USA
sadeh@cs.cmu.edu

Jason Hong
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213 USA
jasonh@cs.cmu.edu

CHI 2010, April 10–15, 2010. Atlanta, Georgia, USA
ACM 978-1-60558-930-5/10/04.

Abstract
Default privacy policies have a significant impact on the overall dynamics and success of online social networks, as users tend to keep their initial privacy policies. In this work-in-progress, we present a new method for suggesting privacy policies for new users by exploring knowledge of existing policies. The defaults generation process performs a collaborative analysis of the policies, finding personalized and representative suggestions. We show how the process can be extended to a wide range of domains, and present results based on 543 privacy policies obtained from a live location-based social network. Finally, we present a user interaction model that lets the user retain control over the default policies, allowing the user to make knowledgeable decisions regarding which default policy to take.

Keywords
information disclosure, privacy, default policies, online social networks, location sharing technology

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: user-centered design
H.5.3 Group and Organization Interfaces evaluation collaborative computing

Introduction

Past work has found that default policies can have a profound impact on users' final policies and their overall use of a system. For example, users tend not to change default calendar sharing settings [7], online social network privacy settings [2, 1, 5], and even organ donation choices [4]. The reasons behind this conservative approach are not fully explored, but there is some evidence that they go beyond the user burden involved in changing the defaults. For example, an experimental study by McKenzie et al. [6] shows that defaults are considered as authoritative recommendations by users.

In this report, we present **collaborative policy analysis**, a method for finding flexible and configurable default privacy policies in online social networks. The defaults are generated using a machine-learning approach, which analyzes large number of existing policies in order to find representative policies that have high likelihood of being relevant to new users. Very much like a recommendation system [3], our method provides new users a selection of policies that similar people have previously selected in the system.

Our user interaction model allows users to understand and configure the default policies through a wizard interface. When a new user first interacts with our system, she is presented with a limited set of personalized default policies, which she can choose from and configure. For example, in a location-based social network, a new user who is a university student will receive several options for a policy that allows other university students to locate her: allow students to locate all the time, allow students to locate only when she is on campus, or do not allow students to locate at all. The user can choose not to accept any of the suggestions, or to adapt each of the suggestions for her needs.

In a previous study [8], we were able to learn default policies obtained from 30 users, who gave detailed feedback on possible executions of their privacy policy. Collaborative policy analysis is similarly motivated, with several improvements: it can be used to analyze existing policies, it offers personalize suggestions, and it can be extended to a wide range of policy types. The last point is particularly noteworthy. Collaborative policy analysis uses generic similarity measures in order to handle policies from a wide range of domains, including social networks, role-based access control, parental controls, and firewall configuration.

We evaluate our method using 543 privacy policies by individual users obtained from Locaccino, a live location-based social network [9]. In this paper, we give examples of the type of policies that are generated by the analysis algorithm, and how the characteristics of the existing policies impact the suggestions. In our future research, we plan to evaluate our method by comparing it to opt-in and opt-out default policies, and to study how default policies impacts the dynamics of the social network.

Collaborative Policy Analysis

In the collaborative policy analysis algorithm, we search for a distinctive set of representative suggestions from the current set of existing policies. A "representative" policy is similar to a large number of policies in the repository, making it more probable to be similar to the policy new random user would create. In order to provide the user with a diverse set of policies to choose from, we look for several representative policies from distinct groups of policies, which are clustered together according to their characteristics. The collaborative analysis process is comprised of three steps:

1. Calculating similarity between policies.
2. Clustering policies.

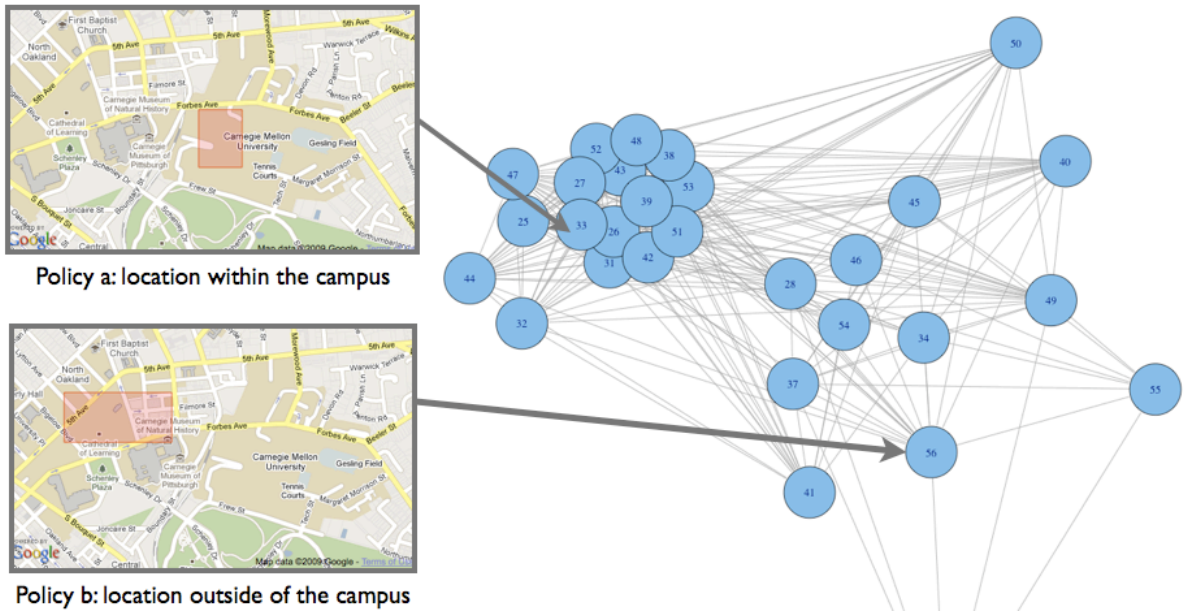

Figure 1: The graph of location sharing policies. Nodes represent policies and edges represent similarity where the proximity indicates strong similarity. The tightly similar policies in the top left contain policies that share location where the user is within the university campus (policy (a), for example). Other policies, e.g., policy (b), share location under different and diverse conditions, which are less popular among users. As a result, the in-campus policies cluster is substantially larger than other clusters.

3. Personalizing policy selection.

In the first step of the collaborative policy analysis algorithm, we measure the similarity between every pair of policies. Similarity is determined by comparing the different properties of the policy. The algorithm is designed to analyze diverse types of policies, comparing any two policies as long as the policy properties can be defined as logic-based sets.

Let us imagine a simple policy in mobile social network that allows access only when the user is within a predefined geographical area. When comparing two policies, we divide the areas defined by the policies to a set of discrete area polygons. We calculate the similarity as the overlap between the areas (formally put, the ratio between the polygons in the intersection and the polygons in the union). A policy that discloses the user's location within the university campus is more similar to a policy that discloses the location within the university's neighborhood than a policy that discloses the lo-

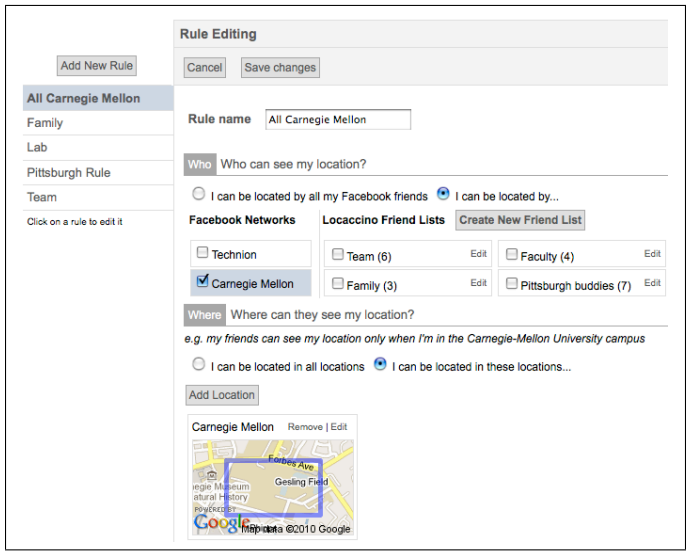

Figure 2: The privacy settings page. Users can create rules for specific friends and social groups, define the location and time in which they want their location to be shared.

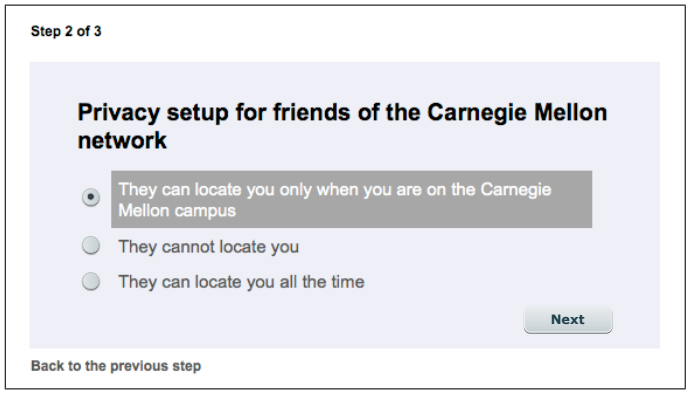

Figure 3: The default privacy settings wizard. In this stage of the wizard, the user can choose between several policies regarding sharing location with the university's students.

cation within the whole city. This similarity measure, called the Jaccard set similarity coefficient, can be used to find the similarity between a wide range of policy properties as long as they can be represented as sets. These include social groups, location restrictions, time restrictions,types of information controlled by the policy, IP ranges in firewall configuration, files in a file system and so fourth.

In the second step, we divide the policies into distinct clusters in order to find a varied set of distinct policies. Figure 1 depicts the results of the applying the k-means algorithm to a part of our policy database, where policies have location, time and social-group properties. A cluster of policies which allow university students to locate the user within the campus borders is visible in the left-top corner of the figure. We rank the policies within each cluster according to their the number of policies they are similar to and the amount of similarity.

In the last step, we personalize the suggestions by selecting policies only from the clusters which are relevant to the user. We compare the information we have regarding the new user with the information we have regarding the users in each cluster, and retrieve policies only form the appropriate clusters. For example, from clusters of policies by users from the same Facebook network the user belongs to. The top ranking

policies from the relevant clusters are then presented to the user.

We evaluated our method in the context of Locaccino, a mobile location social network. Locaccino is based on users' existing social networks on Facebook, allowing users to share their location using software installed on their laptops and mobile phones. Users can define expressive location disclosure rules which determine the exact circumstances under which location information is disclosed. Each rule is comprised of three parts: the groups of Facebook friends that can see their location (groups are manually selected or based on Facebook Networks, e.g. "Georgia Tech"); time restrictions in which location is shared (e.g., weekdays between 9 am - 5 pm); and location restrictions, which define geographic areas in which location is shared.

Our analysis was based on 543 different policies, each created by a distinct user. Users were recruited from the university population during the course of several studies, or have been invited to use the system by study participants. Users were using the system for periods ranging from two days and several months (the median time is 20 days). Users who did not have any friends in the system were omitted from this report. As an example of the algorithm's outcome, the policies for sharing the location with university students are:

- To share location only when the user is on Carnegie-Mellon campus.
- To deny all requests.
- To share location at all times.

User Interaction Model

Our user interaction model lets the user retain full control over the default policies, while keeping the number of suggestions comprehensible and usable. The user is presented with the default policies suggestions using a wizard, where the user can select a rule for the primary dimension of the policy in each step. The primary dimension can be configured according to the policy's domain. In the case of social network privacy settings, the primary dimension is the social group for which the user defines access settings.

Figure 3 depicts a screenshot of one of the steps of the wizard, in which users are asked to choose between several options regarding sharing their location with university friends. The options displayed in the wizard are drawn from the top ranking policies in the clusters which are relevant to the given user. In our example, the user is a student in Carnegie Mellon university, and as a result, receives options from clusters by users that have the same affiliation. The options are ordered according to their ranking. The user can always opt-out from choosing a particular default and configure all or part of the policy manually.

Conclusions and Future Work

In this work-in-progress, we presented a method for suggesting configurable privacy policy defaults by analyzing existing policies. We gave a high-level overview on how machine learning techniques can be used to find default policy suggestions which may have higher probability of being useful to the user. Along with the algorithmic side of the method, we presented a user interaction model that help the user make knowledgeable selections of the default policies. Our current results have several significant limitations. Mainly, the algorithm and user interface were not evaluated. We plan to conduct lab studies and focus groups in order to evaluate different approaches towards displaying and configuring the defaults.

While users do not often change their initial default settings, we witness a slow pacing change in the policies over time. For example, comparing long term changes in privacy settings on Facebook between 2005 [2] and 2008 [5] shows that users' privacy settings have changed considerably, while the default settings remained the same. Lewis et al. [5] argue that social influence has played a major role in the way users set their privacy settings. In our future research, we will empirically examine the relation between default policies and social dynamics, testing how default policies impact the behavior of users and the dynamics of a social network.

Acknowledgment

This work is supported by NSF User-Controllable Policy Learning grant CNS-0905562, NSF Cyber Trust grant CNS-0627513 and ARO research grant DAAD19-02-1-0389 to Carnegie Mellon Universitys CyLab. Additional support has been provided by Microsoft through the Carnegie Mellon Center for Computational Thinking, FCT through the CMU / Portugal Information and Communication Technologies Institute, and through grants from France Telecom and Nokia.

References

[1] Bonneau, J., and Preibusch, S. The privacy jungle: On the market for data protection in social networks. In *The Eighth Workshop on the Economics of Information Security* (2009).

[2] Gross, R., Acquisti, A., and Heinz, H. J. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society* (2005).

[3] Hill, W., Stead, L., Rosenstein, M., and Furnas, G. Recommending and evaluating choices in a virtual community of use. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems* (1995), pp. 194–201.

[4] Johnson, E. J., and Goldstein, D. G. Do defaults save lives? *Science 302* (2003), 1338–1339.

[5] Lewis, K., Kaufman, J., and Christakis, N. The taste for privacy: An analysis of college student privacy settings in an online social network. *Computer-Mediated Communication 14*, 1 (2008).

[6] McKenzie, C. R., Liersch, M. J., and Finkelstein, S. R. Recommendations implicit in policy defaults. *Psychological Science* (2006).

[7] Palen, L. Social, individual and technological issues for groupware calendar systems. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems* (1999), pp. 17–24.

[8] Ravichandran, R., Benisch, M., Kelley, P. G., and Sadeh, N. M. Capturing social networking privacy preferences: Can default policies help alleviate tradeoffs between expressiveness and user burden? In *Privacy Enhancing Technologies* (2009), pp. 1–18.

[9] Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M., and Rao, J. Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing 13*, 6 (August 2008), 401–412.