

Intelligent Content-based Privacy Assistant for Facebook

Michal Jakob, Zbyněk Moler, Michal Pěchouček
Agent Technology Center, Dept. of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Praha, Czech Republic
{jakob, moler, pechoucek}@agents.fel.cvut.cz

Roman Vaculín
IBM T.J. Watson Research Center
Hawthorne, NY 10532 USA
vaculin@us.ibm.com

Abstract—Although most online social networks now offer fine-grained controls of information sharing, these are rarely used, both because their use imposes additional burden on the user and because there are too many control settings for an average user to handle. To mitigate this problem, we have developed an Intelligent Privacy Assistant for Facebook that partially automates the assignment of sharing permissions, taking into account the content of the information published and user’s high-level sharing policies. The Assistant uses a novel social web privacy language, employs named entity recognition algorithms to annotate sensitive parts of published information and an answer set programming system to evaluate user’s privacy policies and determine the list of safe recipients. On a test scenario, the Assistant reached 73.8% and 95.2% performance in correctly determining safe and unsafe recipients, respectively.

Keywords-privacy protection, social web, information extraction, policies, Facebook

I. INTRODUCTION

With the increasing volume of information shared on online social networks, maintaining user’s privacy becomes a major concern. A barrier to efficient privacy management is the complexity of the mechanisms through which information propagation in social networks is controlled. E.g. Facebook, the largest online social network, has tens of privacy settings controlling the access to user information. Though numerous, the controls are simplistic and have minimum expressive power – the user can only restrict access based on the datatype of the information published and/or the social distance to the recipient (i.e. friends, friends of friends etc.). In reality, however, the decision on who should/should not view a particular piece of information is greatly influenced by the *content* of the information published, and the attributes of its potential recipients and their position within the social network of the publisher.

In order to address these problems, we have developed algorithms a prototype Facebook application – termed *Intelligent Privacy Assistant* – that extends the sharing control to consider a much wider range of factors – including information content and social context – while reducing the need for manual privacy control on a case-by-case basis. Although the content-aware approach is not entirely new and has been employed to control privacy in e-mail conversations

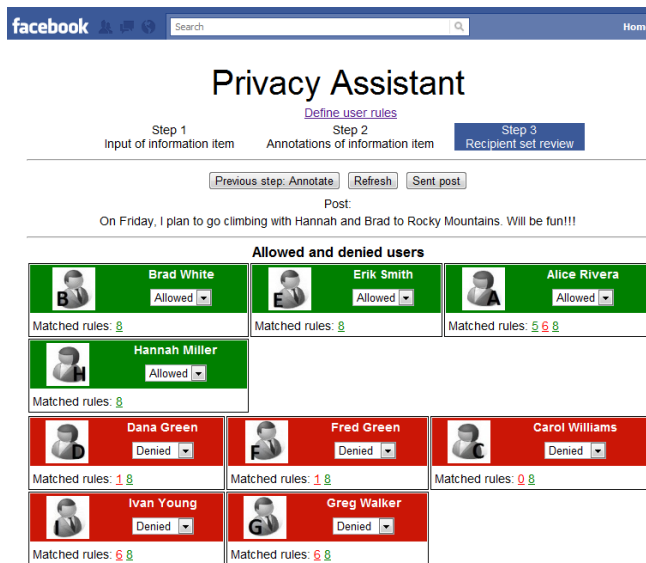


Figure 1. Output of the recommendation step of the Privacy Assistant showing the allowed (top/green) and denied (bottom/red) recipients for a user’s post as determined by the user’s policies.

[1] in the past, our application is the first which employs the approach in the context of social networks.

II. FACEBOOK PRIVACY ASSISTANT

The Privacy Assistant allows the user to define sharing policies for automatically selecting users that can/cannot see his published information. From the user’s perspective, the Assistant consists of two parts. The *sharing policy editor* allows the user to define sharing policies using a newly proposed Social Web Privacy Language (see below). Sharing policies are defined in terms of rules which, in contrast to existing policy languages, can also reference the content of the information and the user’s social graph, which is automatically extracted from Facebook.

Once the sharing policies are defined, the user uses the *managed status publisher* for assisted information sharing. The managed status publisher replaces the standard Facebook status update form. Its usage consists of three steps – in the first step, the user fills in the status update he wants to publish. In the second step, the user is presented with a set of automatically extracted *privacy annotations*

which highlight potentially sensitive parts of the input text (in particular references to people, places, times/dates and activities). The user can review and manually modify the proposed annotations. Once the annotations are confirmed, the Assistant evaluates user’s sharing policies and proposes a set of recipients that should be allowed to see the published information (see Figure 1). Again, the user can review and modify the proposed recipient set. Once confirmed, the information is posted on Facebook with correctly assigned sharing permissions. After the user has fine-tuned his rules, the review steps can be skipped, resulting in fully automated information sharing with content- and context-specific recipient assignment. A demo video of the application is available at project’s website¹.

III. KEY BUILDING BLOCKS

The implementation of the Privacy Assistant is based on three fundamental building blocks. More details can be found in [2].

A. Social Web Privacy Language

Because existing privacy languages do not support content- or social-context-related concepts, we have introduced a new *Social Web Privacy Language* (SWPL). In addition to basic concepts taken from existing privacy policy languages (e.g. [3]), SWPL provides a vocabulary for describing social networks. To avoid pitfalls of general natural language processing and to allow efficient reasoning, we limit our attention to those content elements that have a high chance of implying sensitivity of the content. SWPL therefore provides a set of privacy annotation predicates based on the concept of *Five W’s* (*who, what, when, where, why*)², which is employed in journalism and police investigation as a basis of information gathering.

B. Automated Privacy Annotation

We employ automated named entity recognition (NER) algorithms, a particular type of information extraction techniques, for automatically identifying parts of the input text that should be assigned privacy annotations. To support full range of annotation types and to reduce annotation errors, we use ensemble classification techniques to aggregate outputs of multiple NER methods to a single, more reliable output.

The integration of individual NER algorithms uses the *GATE framework* [4] and the following specific NER systems: Name Finder from OpenNLP, Stanford NER, NER from LingPipe and Gazetteer from the GATE framework.

C. Sharing Policy Evaluation

We employ the *DLV reasoner*³, based on disjunctive dialog with constraints, true negation and complex queries, for

the evaluation of sharing policies. It has been chosen because of its support for non-monotonic reasoning, arithmetics, recursive predicates, and reasoning with very large datasets.

IV. EVALUATION

Unfortunately, there is currently no dataset with Facebook posts and information on the underlying social network. We therefore created a dataset based on a realistic scenario and samples of real Facebook posts⁴, comprising 10 users arranged in several groups of different social relationships (family, work, romance) and 110 status updates with different topics and sensitivity levels. On this dataset, the implemented Privacy Assistant achieved 73.8% and 95.2% performance in correctly denying and allowing, respectively, recipients access to the published information. We are currently working on expanding the dataset.

V. CONCLUSION

We have implemented a Facebook application that assists in safely sharing potentially sensitive information in online social networks. By partially automating user’s sharing decisions, the application aims to make selective sharing easier for a typical Facebook user and increases chances that a user will be able to properly manage her online privacy. As the next step, we plan to incorporate methods for automatically learning user sharing policies from her past sharing decisions in order to limit the need for explicit policy specification, which is difficult for some users.

ACKNOWLEDGMENT

Supported by a Google Research Award and by the Czech Ministry of Education, Youth and Sports under grant "Decision Making and Control for Manufacturing III" (grant no. MSM 6840770038).

REFERENCES

- [1] N. Boufaden, W. Elazmeh, Y. Ma, S. Matwin, N. El-Kadri, and N. Japkowicz, "Peep- an information extraction base approach for privacy protection in email." in *Proceedings of the 2005 Conference on E-mail and Anti-spam*, 2005.
- [2] M. Jakob, Z. Moler, R. Vaculín, and M. Pěchouček, "Content-based privacy management on the social web," in *Web Intelligence for Information Security Workshop of WI-IAT 2011*, 2011.
- [3] P. Ashley, S. Hada, G. Karjoth, C. Powers, and M. Schunter, "Enterprise privacy authorization language (EPAL 1.2)," W3C, Tech. Rep., 2003.
- [4] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, "GATE: A framework and graphical development environment for robust NLP tools and applications." in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA*, 2002.

¹<http://agents.fel.cvut.cz/projects/privacy20>

²see e.g. http://en.wikipedia.org/wiki/Five_Ws

³<http://www.dbai.tuwien.ac.at/proj/dlv/>

⁴<http://www.facebook.com/group.php?gid=51097603092&v=wall>