Exploring the Viability of Tie Strength and Tags in Access Controls for Photo Sharing

Ricard L. Fogues
Universitat Politecnica de Valencia
Cami de Vera s/n
Valencia, Spain
rilopez@dsic.upv.es

Jose M. Such
King's College
Department of Informatics
London, UK
jose.such@kcl.ac.uk

Agustin Espinosa
Universitat Politecnica de Valencia
Cami de Vera s/n
Valencia, Spain
aespinosa@dsic.upv.es

Ana Garcia-Fornes
Universitat Politecnica de Valencia
Cami de Vera s/n
Valencia, Spain
agarcia@dsic.upv.es

ABSTRACT

Social Network Sites (SNSs) such as Facebook and Google+ allow users to store and share large collections of photos. SNSs offer access controls that protect those photos from unwanted audiences. However, due to the lack of usability of these access controls, people struggle to configure them. First, we collected sharing policies for photos in a study with 34 Facebook users. Then, we define three metrics that enable researchers to evaluate the ease of use and complexity of access controls for photo sharing, and, employing the data collected in the study, we evaluate 15 access controls, each one with a different combination of attributes. The results obtained show that an access control that takes into account groups, tags, and the tie strength of relationships can be managed more easily than current approaches, reducing the burden of configuring the privacy settings for photos on SNSs.

CCS Concepts

•Security and privacy → Social aspects of security and privacy;

Keywords

Privacy; social media; access controls

1. INTRODUCTION

Photos are one of the most common items uploaded to SNSs. People share photos on SNSs because they perceive a great pay-off in terms of friendship and other social opportunities [4]. However, since photos can disclose sensitive information, they require complex privacy settings [1]. In order to control how photos are disclosed, SNSs offer basic access controls [2]. These access controls are based on groups of contacts. For example, Facebook has friend lists and Google+ offers friend circles. Users can employ these groups of contacts to specify who is allowed to access a piece of content.

Although access controls of SNSs aim at simplicity, users struggle to manage and understand them [9]. To improve the understandability and usability of access controls, research works propose adding different attributes to access controls in SNSs [2, 3]. These attributes include variables such as social distance, tie strength, or groups of contacts. However, few works evaluate what combinations of attributes can potentially cover the privacy preferences of the users with the lowest possible complexity. In this study we evaluate the effects of the addition of new attributes (and combinations of them) to photo access controls. Our contribution is twofold: we (1) propose three quantitative metrics that enable us to evaluate the impact of new attributes on current access controls.; and (2) employing the proposed metrics, we evaluate the performance of 15 combinations of new and current attributes for access controls.

2. RELATED WORK

Yeung et al. [12] prototyped the management of privacy for photos that considers content type. Hart et al. [6] proposed a mechanism to manage privacy for blogs based on tags. Their mechanism enables users to define groups manually or to group potential viewers by attributes that they all share (e.g., workplace or same school). The main focus of their study is to compare basic sharing policy mechanisms for blogs with a tag-based approach. The authors did not use real data from the participants, instead, they created artificial data for imaginary users and asked the participants to manage that data as if it was theirs. Thus, they do not examine users' actual preferences, as we do. Their results show that an approach that uses tags is more usable than one that does not. However, they do not evaluate if other combinations of attributes can further improve the performance. Squicciarini et al. [11] proposed a sharing policy recommender tool which considers tags and contact groups.

SAC'17, April 3-7, 2017, Marrakesh, Morocco

http://dx.doi.org/xx.xxxx/xxxxxxx.xxxxxxx

Since they aimed at accurately inferring new sharing policies, they did not evaluate the effect of tags as a new attribute for an access control. Further, they evaluated the performance of their tool using predefined photos, instead, we use personal photos from the participants. Klemperer et al. [7] evaluated the usability of an access control based exclusively on tags. The authors aimed at evaluating whether tags can be used to organize photos and define their privacy at the same time. However, they did not compare the performance of their access control with SNSs' current approaches, or with other access controls with different attributes.

3. ATTRIBUTES OF ACCESS CONTROLS

The attributes evaluated in our study are the following:

Tags (*Tag*): The different categories that define the content of the photo.

Communities (*Com*): Groups of contacts created by the users.

Tie Strength (*Tie*): The individual tie strength that users have with each one of their contacts based on a Likert scale from 1 to 5 (1 = weak tie, 5 = strong tie).

Individual Contacts (*Ind*): The current access control of Facebook allows users to specify individual contacts in sharing policies. We also consider this attribute for the different access controls that we evaluate in this section.

We use a combination of abbreviations of the attributes to name an access control. For example, the name of an access control that takes into account tags and tie strength is *TagTie*.

4. METHOD

Our investigation is based on real data retrieved from Facebook users. The information that we needed for our study is divided into three types: (i) the characteristics of the relationships between the participants and their contacts, (ii) the sharing policies that the participants apply to the photos on their Facebook accounts, and (iii) the tags of these photos.

4.1 Participants

The sample consisted of 11 women and 22 men. The mean age of the participants was 25.66 ($SD = 6.19$) with 18 years as the minimum age and 45 years as the maximum age. For education level, the majority of them had a college degree (20), 6 of them had a PhD, another 6 of participants had a high school degree, and one participant had a primary school degree. Finally, 76% of the participants were students and the other 24% were working.

4.2 Collection of Social Data

To collect the information of the relationships of the participants, we used the BFF application [5]. This is a Facebook application that helps users organize their relationships. BFF automatizes the process of friend grouping and tie strength definition. BFF collects predictive variables from the user's profile. These variables include data such as the number of messages exchanged with a friend, who appears in the photos of the user, or the total number of friends. With the collected data, BFF infers tie strength values and friend communities. BFF represents tie strength on a Likert scale 1–5 (1 = minimum, 5 = maximum). The results yielded by BFF can be refined by the users. Communities were not exclusive, one single contact could be included in several communities if the participant considered it appropriate. In total we collected 735 communities and the average number of communities per participant was 22.27.

4.3 Definition of Sharing Policies and Photo Tags

The participants defined sharing policies for their photos on Faceboook. First, our application collected the photos of the participants from their Facebook profiles. The photos were sorted and organized using the same album structure that the users have on their profiles. Since the access control of Facebook is based on individual contacts and contact groups, during this step, the participants defined sharing policies using the same groups and individuals that they corrected or created during the previous part of the study. Participants were asked to define their ideal policies for individual photos, and they created as many as they found necessary. The application we built enabled participants to assign the same policy to every photo in the same album if they found that appropriate. As in Facebook, participants were told that blocking takes precedence over granting access.

A majority of photos (64.89%) were assigned a public policy. On average, each photo was accessible by 89.04% of contacts. However, if we analyze only the photos that did not have a public policy we observe that the defined sharing policies were somewhat restrictive. On average, 54.11% of contacts could access every photo with a non-public policy. We found that 4 male participants used only public policies for their photos. Since this study is focused on how users define sharing policies, we do not consider their information for the evaluation of the access controls.

Participants were required to classify each album with one or more tags after defining their ideal sharing policies. The predefined tags were: family, close friends, colleagues, party, kids, travel, animals, self-portrait, fun, artistic, and other. This set of tags was extracted from the most common[1] tags used in Flickr the popular photo-exchange social network[2]. Overall, 9% of the photos were classified with *other* as one of their tags. However, only 1.13% of the photos were classified exclusively as *other*, thus, the predefined tags covered almost the entire set of photos of all participants. We told the participants that they had to assign tags for the content of the photos, not for their appropriate audience. To minimize the risk of participants using tags to define audiences, we deliberately set the tagging task after the sharing policy definition task.

5. EVALUATION

In total, we defined 15 different access controls using the different combinations of the attributes explained above. We aim at comparing the performance of these 15 access controls. However, asking the participants to define their privacy preferences with 15 access controls would not have been feasible in terms of time and task complexity. Therefore, to find how the policies defined by the participants when using each access control model might look like, we use decision-tree classifiers. These classifiers automatically generate a representation (a tree of rules) of how users could use the available attributes in each access control model to define sharing policies that match their privacy preferences. Since the rules gen-

[1] Even though the tag *colleagues* is not a popular tag in Flickr, we added it because we felt it makes sense in a friendship-focused social network such as Facebook.

[2] http://www.flickr.com/photos/tags/

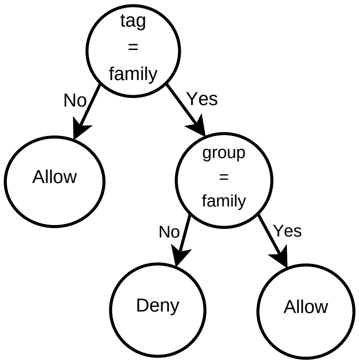

Figure 1: Example tree.

erated are, somewhat, simulations of participants' general privacy preferences, they do not capture every detail. For example, a participant may share every family photo with his father except a photo that depicts the preparation of a surprise birthday party for his father. Our method is not able of capturing this nuance, and the birthday preparation photo will be incorrectly classified (his father will be allowed to see it).

For the creation and evaluation of the decision-tree classifiers we used the C4.5 decision tree [10] and its implementation in the Weka[3] data mining tool. All the classifiers considered two classes: allow or deny. For each photo and contact, the classifier had to decide if the photo could be accessed by that contact (allow) or not (deny). The number of dimensions of the feature vector of each classifier depended on the attributes considered by the access control. For example, if the access control takes into account tags, communities, and tie strength, then, the vector of features will have a dimension for each possible photo tag, each community defined by the participant, and one dimension for the tie strength.

5.1 Metrics

Coverage: This measures how well the privacy preferences are correctly represented by the automatically created rules. In other words, for each pair (photo and contact) we checked whether or not the tree generated for that specific access control classified correctly the pair into one of the two classes (allow or deny). The coverage gives us information about how well the access control can express the actual privacy preferences of the user.

Number of rules: To find an appropriate access control, it is necessary to consider the trade-off between coverage and complexity. If an access control requires many rules to cover the privacy preferences of the user, it is likely that the access control will have a low usability. To measure the number of rules generated by each access control, we counted the number of leaves in the generated tree. For example, Figure 1 shows a simple example tree. In this tree, there are three leaves, thus, the access control has three rules: (i) family members can see family photos; (ii) non-family members cannot see family photos; and (iii) non-family photos are public.

Complexity level: Similar to the number of rules, rule complexity can be detrimental to the usability of the access control. The level of complexity of a rule is given by the number of tags, community identifiers, tie strength thresholds, and individual contact identifiers utilized. In the example shown in Figure 1, there are two rules with complexity 2 (the community and tag attributes are used), and one rule with a complexity of 1 (only the tag attribute is used).

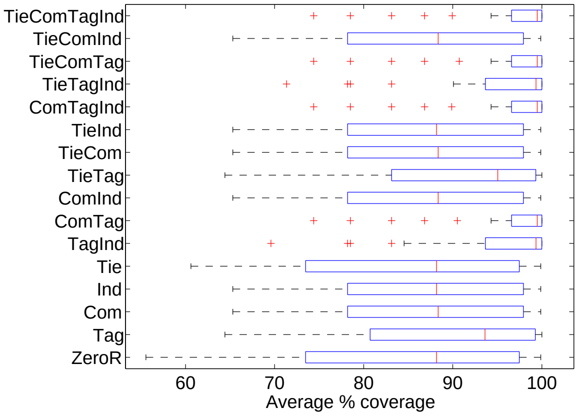

Figure 2: Coverage of access controls created with different attribute combinations.

6. RESULTS

Figure 2 shows the coverage obtained by the different combinations of attributes. We use the coverage obtained by the*ZeroR* classification method as a benchmark for the different access controls. This method is the simplest classification method since it relies on the target and ignores all predictors. The *ZeroR* classifier simply predicts the majority class.

According to a Lilliefors test [8] with a 95% confidence interval, the coverage values obtained by the different access controls come from a normally distributed population. Therefore, to test whether or not the differences in coverage were significant, we performed a series of t-tests with a 95% confidence interval. Since the data used for all the access controls was the same, we used paired t-tests. To counteract the problem of multiple comparisons, we applied Holm-Bonferroni correction to the series of t-tests. The test shows that some differences are not statistically significant. The statistical differences show four different groups of access controls. Specifically, the group of access controls with the highest coverage is formed by: TieComTagInd, TieComTag, ComTagInd, ComTag, and TagInd. The second group is formed solely by TieTagInd. The third group contains TieTag and Tag. Finally, the fourth group is formed by the rest of access controls: TieComInd, TieInd, TieCom, ComInd, Tie, Ind, and Com. Overall, according to these results, tags improves the coverage of an access control the most, followed by communities, individuals, and tie strength, which affects coverage the least.

Figure 3 shows the number of rules generated by each access control. The results show that there were big differences in the number of rules generated. Several access controls require a number of rules that make them unmanageable for a human user. The access control with the least number of rules was *Tie*. Actually, this low number of rules explains its poor performance; the trees generated with this access control were almost the same as the rules generated by the *ZeroR* classifier. Among the access controls with good performance, *ComTag* has a slightly lower average number of rules.

Figure 4 depicts the complexity levels of the rules generated by the access controls. In general, the median level of complexity was around 4. Obviously, the lack of expressivity of *Tie* limited the complexity of this access control. *Ind*, *TieInd*, *TieTagInd* and

[3] http://www.cs.waikato.ac.nz/ml/weka/

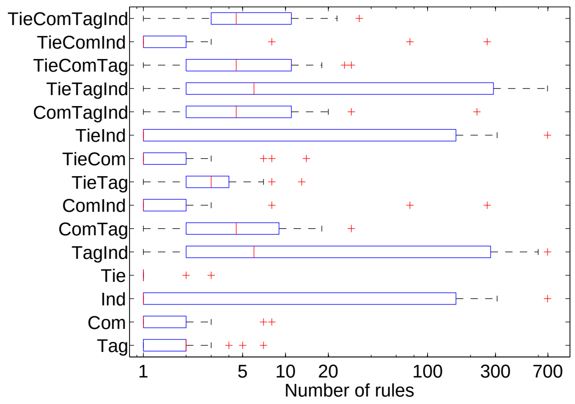

Figure 3: Number of rules generated by each access control.

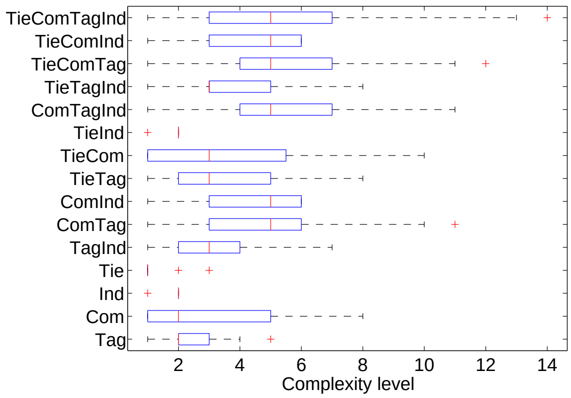

Figure 4: Complexity level of the rules generated by each access control.

TagInd produced also rules with a low level of complexity; however, as shown in Figure 3, they generate a large number of rules. Overall, *ComTag*, *ComTagInd*, *TieComTag*, and *TieComTagInd* offered the most balanced results: good coverage, a small number of rules, and low levels of complexity.

7. DISCUSSION AND CONCLUSIONS

We propose 15 different acces controls and three metrics to evaluate them. Using real privacy preferences, we compare the performance of these 15 access controls. According to our results, an access control that takes into account tags, communities, and tie strength requires a low number of rules with low complexity to express the general privacy preferences of the users with coverage that is good enough.

Analyzing the two new attributes individually, on the one hand, the results obtained in our study point out that access controls with the attribute tags achieve good coverage with low complexity. This shows that tags play a key role during sharing policy definition.

On the other hand, the tie strength attribute does not show an impact on access controls as positive as tags. One of the reasons behind this could be that participants did not assign tie strength values depending on how much they share on the SNS but outside of it. Future work should investigate whether users employing an access control that uses tie strength as a means to define sharing policies create less complex and more accurate policies than those who use an access control without it.

8. ACKNOWLEDGMENTS

We thank Ministerio de Economía y Competitividad (grant TIN2014-55206-R) for partial support.

9. REFERENCES

[1] A. Besmer and H. Richter Lipford. Moving beyond untagging: photo privacy in a tagged world. In *Proc. CHI*, pages 1563–1572, 2010.

[2] B. Carminati, E. Ferrari, R. Heatherly, M. Kantarcioglu, and B. Thuraisingham. Semantic web-based social network access control. *Computers & Security*, 30(2âĂŞ3):108 – 115, 2011. Special Issue on Access Control Methods and Technologies.

[3] Y. Cheng, J. Park, and R. Sandhu. A user-to-user relationship-based access control model for online social networks. *Data and Applications Security and Privacy XXVI*, pages 8–24, 2012.

[4] N. Ellison, C. Steinfield, and C. Lampe. The benefits of facebook friends: Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168, 2007.

[5] R. Fogues, J. Such, A. Espinosa, and A. Garcia-Fornes. Bff: A tool for eliciting tie strength and user communities in social networking services. *Information Systems Frontiers*, 16(2):225–237, 2014.

[6] M. Hart, C. Castille, R. Johnson, and A. Stent. Usable privacy controls for blogs. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, pages 401–408. IEEE, 2009.

[7] P. Klemperer, Y. Liang, M. Mazurek, M. Sleeper, B. Ur, L. Bauer, L. F. Cranor, N. Gupta, and M. Reiter. Tag, you can see it!: using tags for access control in photo sharing. In *Proc. CHI*, pages 377–386, 2012.

[8] H. W. Lilliefors. On the kolmogorov-smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association*, 64(325):387–389, 1969.

[9] H. Lipford, A. Besmer, and J. Watson. Understanding privacy settings in facebook with an audience view. In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, pages 1–8. USENIX Association Berkeley, CA, USA, 2008.

[10] J. R. Quinlan. *C4.5: programs for machine learning*, volume 1. Morgan kaufmann, 1993.

[11] A. Squicciarini, S. Sundareswaran, D. Lin, and J. Wede. A3p: adaptive policy prediction for shared images over popular content sharing sites. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, pages 261–270. ACM, 2011.

[12] C. Yeung, L. Kagal, N. Gibbins, and N. Shadbolt. Providing access control to online photo albums based on tags and linked data. In *Proceedings of the AAAI Spring Symposium on Social Semantic Web: Where Web*, volume 2, 2009.