SURVEY ON SECURE AND TIME CONFINED IMAGE SHARING ON WEBSITES

[1]Rajani Kongala and [2]G.V.Ramana, [3]J.V Krishna
[1]Department of Computer Science and Engineering , Sree Vahini Institute of Science and Technology,Tiruvuru.,A.P.,India.
[2,3] Associate Professor ,Sree Vahini Institute of Science and Technology,Tiruvuru.,A.P.,India.

ABSTRACT: With the increasing volume of images users share through social sites, maintaining privacy has become a major problem. In light of these incidents, the need of tools to help users control access to their shared content is highly essential. Toward addressing this need, we propose an Adaptive Privacy Policy Prediction (A3P) system to help users compose privacy settings for their images. We examine the role of social context, image content, and metadata as possible indicators of users' privacy preferences. We propose a two-level framework which according to the user's available history on the site determines the best available privacy policy for the user's images being uploaded. Our solution relies on an image classification framework for image categories which may be associated with similar policies, and on a policy prediction algorithm to automatically generate a policy for each newly uploaded image, also according to users' social features.

I.INTRODUCTION

Images are now one of the key enablers of users' connectivity. Sharing takes place both among previously established groups of known people or social circles (e.g., Google+, Flicker or Picasa), and also increasingly with people outside the users social circles, for purposes of social discovery- to help them identify new peers and learn about peers interests and social surroundings. However, semantically rich images may reveal content sensitive information []. Consider a photo of a student's 2012 graduation ceremony, for example. It could be shared within a Google+ circle or Flicker group, but may unnecessarily expose the student's family members and other friends. Sharing images within online content sharing sites, therefore, may quickly lead to unwanted disclosure and privacy violations [3], [25]. Further, the persistent nature of online media makes it possible for other users to collect rich aggregated information about the owner of the published content and the subjects in the published content [3], [21], [25]. The aggregated information can result in unexpected exposure of one's social environment and lead to abuse of one's personal information. Most content sharing websites allow users to enter their privacy preferences. Unfortunately, recent studies have shown that users struggle to set up and maintain such privacy settings [1], [12], [23], [34]. One of the main reasons provided is that given the amount of shared information this process can be tedious and error-prone. Therefore, many have acknowledged the need of policy recommendation systems which can assist users to easily and properly configure privacy settings [7], [23], [29], [31]. However, existing proposals for automating privacy settings appear to be inadequate to address the unique privacy needs of images [3], [5], [42], due to the amount of information implicitly carried within images, and their relevance with respect to the online social environment wherein they are exposed.

II.RELATED WORK

Our work is related to works on privacy setting configuration in social sites, recommendation systems, and privacy analysis of online images. Privacy Setting Configuration Several recent works have studied how to automate the task of privacy settings (e.g. [7], [16], [21], [23], [28], [29]). Bonneau et al. [7] proposed the concept of

privacy suites which recommend to users a suite of privacy settings that "expert" users or other trusted friends have already set, so that normal users can either directly choose a setting or only need to do minor modification. Similarly, Danezis [8] proposed a machine-learning based approach to automatically extract privacy settings from the social context within which the data is produced. Parallel to the work of Danezis, Adu-Oppong et al. [16] develop privacy settings based on a concept of "Social Circles" which consist of clusters of friends formed by partitioning users' friend lists. Ravichandran et al. [31] studied how to predict a user's privacy preferences for location-based data (i.e., share her location or not) based on location and time of day. Fang et al. [29] proposed a privacy wizard to help users grant privileges to their friends. The wizard asks users to first assign privacy labels to selected friends, and then uses this as input to construct a classifier which classifies friends based on their profiles and automatically assign privacy labels to the unlabeled friends. More recently, Klemperer et al. [21] studied whether the keywords and captions with which users tag their photos can be used to help users more intuitively create and maintain access-control policies. Their findings are in line with our approach: tags created for organizational purposes can be repurposed to help create reasonably accurate access-control rules. The aforementioned approaches focus on deriving policy settings for only traits, so they mainly consider social context such as one's friend list. While interesting, they may not be sufficient to address challenges brought by image files for which privacy may vary substantially not just because of social context but also due to the actual image content. As far as images, authors in [42] have presented an expressive language for images uploaded in social sites. This work is complementary to ours as we do not deal with policy expressiveness, but rely on common forms policy specification for our predictive algorithm. In addition, there is a large body of work on image content analysis, for classification and interpretation (e.g., [9], [15], [38], [47]), retrieval ([13], [14] are some examples), and photo ranking [36], [41], also in the context of online photo sharing sites, such as Flickr [11], [30], [37]. Of these works, Zerr's work [44] is probably the closest to ours. Zerr explores privacy aware image classification using a mixed set of features, both content and meta-data. This is however a binary classification (private vs. public), so the classification task is very different than ours. Also, the authors do not deal with the issue of coldstart problem.

III.SYSTEM OVERVIEW

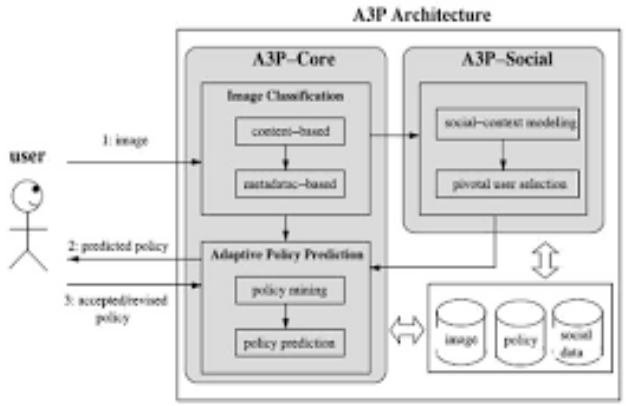
The A3P system consists of two main components: A3Pcore and A3P-social. The overall data flow is the following. When a user uploads an image, the image will be first sent to the A3P-core. The A3P-core classifies the image and determines whether there is a need to invoke the A3P-social. In most cases, the A3P-core predicts policies for the users directly based on their historical behavior. If one of the following two cases is verified true, A3P-core will invoke A3P-social: (i) The user does not have enough data for the type of the uploaded image to conduct policy prediction; (ii) The A3Pcore detects the recent major changes among the user's community about their privacy practices along with user's increase of social networking activities (addition of new friends, new posts on one's profile etc). In above cases, it would be beneficial to report to the user the latest privacy practice of social communities that have similar background as the user. The A3P-social groups users into social communities with similar social context and privacy preferences, and continuously monitors the social groups. When the A3P-social is invoked, it automatically identifies the social group for the user and sends back the information about the group to the A3P-core for policy prediction. At the end, the predicted policy will be displayed to the user. If the user is fully satisfied by the predicted policy, he or she can just accept it. Otherwise, the user can choose to revise the policy.

A3P FRAMEWORK

Preliminary Notions Users can express their privacy preferences about their content disclosure

preferences with their socially connected users via privacy policies. We define privacy policies according to Definition 1. Our policies are inspired by popular content sharing sites (i.e. Facebook, Picasa, Flickr), although the actual implementation depends on the specific content management site structure and implementation. . In the definition, users in S can be represented by their identities, roles (e.g., family, friend, coworkers), or organizations (e.g., non-profit organization, profit organization). D will be the set of images in the user's profile. Each image has a unique ID along with some associated metadata like tags "vacation", "birthday". Images can be further grouped into albums. As for A, we consider four common types of actions: {view, comment, tag, download}. Last, the condition component C specifies when the granted action is effective. C is a Boolean expression on the grantees' attributes like time, location, and age. For better understanding, an example policy is given below. Example 1: Alice would like to allow her friends and coworkers to comment and tag images in the album named "vacation album" and the image named "summer.jpg" before year 2012. Her privacy preferences can be expressed by the following policy: P: [{friend, coworker}, {vacation album, summer.jpg}, {comment, tag}, (date< 2012)].

IV.SYSTEM ARCHITECTURE



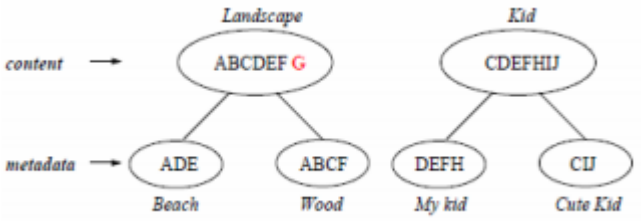
A3P-CORE

There are two major components in A3P-core: (i) Image classification and (ii) Adaptive policy prediction. For each user, his/her images are first classified based on content and metadata. Then, privacy policies of each category of images are analyzed for the policy prediction. Adopting a two-stage approach is more suitable for policy recommendation than applying the common one-stage data mining approaches to mine both image features and policies together. Recall that when a user uploads a new image, the user is waiting for a recommended policy. The two-stage approach allows the system to employ the first stage (i.e., the image classification) to classify the new image and find the candidate sets of images for the subsequent policy recommendation. As for the one stage mining approach, it would not be able to locate the right class of the new image because its classification criteria needs both image features and policies whereas the policies of the new image are not available yet. Moreover, combining both image features and policies into a single classifier would lead to a system which is very dependent to the specific syntax of the policy. If a change in the supported policies were to be introduced, the whole learning model would need to change.

Image Classification

To obtain groups of images that may be associated with similar privacy preferences, we propose a hierarchical image classification which classifies images first based on their contents and then refine each category into subcategories based on their metadata. Images that do not have metadata will be grouped only by content. Such a hierarchical classification gives a higher priority to image content and minimizes the influence of missing tags. Note that it is possible that some images are included in multiple categories as long as they contain the typical content features or metadata of those categories. Moreover, Figure 2 shows an example of image classification for 10 images named as A, B, C, D, E, F, G, H, I, J, respectively. The content-based classification creates two categories: "landscape" and "kid". Images C, D, E and F are included in both categories as they show kids playing outdoor which satisfy the two themes: "landscape" and "kid". These two categories are further divided into subcategories based on tags associated with the images. As a result, we obtain two

subcategories under each theme respectively. Notice that image G is not shown in any subcategory as it does not have any tag; image A shows up in both subcategories because it has tags indicating both "beach" and "wood".



Content-based Classification

Our approach to content-based classification is based on an efficient and yet accurate image similarity approach. Specifically, our classification algorithm compares image signatures defined based on quantified and sanitized version of Haar wavelet transformation. For each image, the wavelet transform encodes frequency and spatial information related to image color, size, invariant transform, shape, texture, symmetry, etc. Then, a small number of coefficients are selected to form the signature of the image. The content similarity among images is then determined by the distance among their image signatures. Our selected similarity criteria include texture, symmetry, shape (radial symmetry and phase congruency [27]), and SIFT [26]. We also account for color and size. We set the system to start from five generic image classes: (a) explicit (e.g., nudity, violence, drinking etc), (b) adults, (c) kids, (d) scenery (e.g., beach, mountains), (e) animals. As a preprocessing step, we populate the five baseline classes by manually assigning to each class a number of images crawled from Google images, resulting in about 1000 images per class. Having a large image dataset beforehand reduces the chance of misclassification. Then, we generate signatures of all the images and store them in the database. Upon adjusting the settings of our content classifier, we conducted some preliminary test to evaluate its accuracy. Precisely, we tested our classifier it against a ground-truth dataset, Image-net.org [18]. In Image-net, over 10 million images are collected and classified according to the wordnet structure. For each image class, we use the first half set of images as the training dataset and classify the next 800 images. The classification result was recorded as correct if the synset's main search term or the direct hypernym is returned as a class. The average accuracy of our classifier is above 94%. Having verified the accuracy of the classifier, we now discuss how it is used in the context of the A3P core. When a user uploads an image, it is handled as an input query image. The signature of the newly uploaded image is compared with the signatures of images in the current image database. To determine the class of the uploaded image, we find its first m closest matches. The class of the uploaded image is then calculated as the class to which majority of the m images belong. If no predominant class is found, a new class is created for the image. Later on, if the predicted policy for this new image turns out correct, the image will be inserted into the corresponding image category in our image database, to help refine future policy prediction. In our current prototype, m is set to 25 which is obtained using a small training dataset.

Adaptive Policy Prediction The policy prediction algorithm provides a predicted policy of a newly uploaded image to the user for his/her reference. More importantly, the predicted policy will reflect the possible changes of a user's privacy concerns. The prediction process consists of three main phases: (i) policy normalization; (ii) policy mining; and (iii) policy prediction. The policy normalization is a simple decomposition process to convert a user policy into a set of atomic rules in which the data (D) component is a single-element set. An example of policy normalization is shown below. Example 2: Consider policy P in Example 1. Suppose that the album "vacation album" contains k images, namely img1 .jpg, img2 .jpg, ..., imgk.jpg. P is normalized into the following set of atomic rules:

Policy Mining

We propose a hierarchical mining approach for policy mining. Our approach leverages association rule mining techniques to discover popular patterns in policies. Policy mining is carried out within the same category of the new image

because images in the same category are more likely under the similar level of privacy protection. The basic idea of the hierarchical mining is to follow a natural order in which a user defines a policy. Given an image, a user usually first decides who can access the image, then thinks about what specific access rights (e.g., view only or download) should be given, and finally refine the access conditions such as setting the expiration date. Correspondingly, the hierarchical mining first look for popular subjects defined by the user, then look for popular actions in the policies containing the popular subjects, and finally for popular conditions in the policies containing both popular subjects and conditions.

Policy Prediction

The policy mining phase may generate several candidate policies while the goal of our system is to return the most promising one to the user. Thus, we present an approach to choose the best candidate policy that follows the user's privacy tendency. To model the user's privacy tendency, we define a notion of strictness level. The strictness level is a quantitative metric that describes how "strict" a policy is. In particular, a strictness level L is an integer with minimum value in zero, wherein the lower the value, the higher the strictness level. It is generated by two metrics: major level (denoted as l) and coverage rate (α), where l is determined by the combination of subject and action in a policy, and α is determined by the system using the condition component. l is obtained via Table 4. In Table 4, all combinations of common subject and common actions are enumerated and assigned an integer value according to the strictness of the corresponding subjects and actions. For example, "view" action is considered more restricted than "tag" action. Given a policy, its l value can be looked up from the table by matching its subject and action. If the policy has multiple subjects or actions and results in multiple l values, we will consider the lowest one. It is worth noting that the table is automatically generated by the system but can be modified by users according to their needs. Then, we introduce the computation of the coverage rate α which is designed to provide fine-grained strictness level. A is a value ranging from 0 to 1 and it will just adjust but not dominate the previously obtained major level. In particular, we define α as the percentage of people in the specified subject category who satisfy the condition in the policy. For example, a user has 5 family members documented in the system and two of them are kids. When he specifies a policy with the condition age > 18, only three family members will satisfy this condition. The corresponding α is then 3/5=0.6. The larger the value of α, the more people are allowed to access the image and hence the policy is less restricted. Therefore, we subtract (1-α) from l to obtain the final strictness level as shown in Equation 2.

L= l-(1-α)

policies, we now need to determine which strictness level fits best to the user's privacy trend. For this purpose, we propose the following approach. We keep monitoring the average strictness level of existing policies in each category of images. The average strictness level is defined as follows:

$$L_{avg} = \frac{\sum_{i=1}^{N_p} L_{p_i}}{N_p}$$

where Lpi denote the strictness level of policy Pi, and Np is the total number of policies Notice that the average strictness level is computed by excluding outlier policies. This is because in some situations, users may define special policies which have a very different strictness level from most of others, either much stricter or much looser. Considering such outliers into the average strictness level calculation would not represent the average case properly. Therefore, when a policy is inserted, we first compare its strictness level with current average strictness level. If the difference is more than a threshold (ξ), we put the policy in the outlier group. In the experiments, we set ξ to 4 because each role of the policy subject has 4 different strictness levels as shown in Table 4. Also, the change on the policy preferences being more than 4 is considered prominent as it exceeds one quarter of the maximum strictness level. As time evolves, the average strictness levels in each category form a curve as shown in Figure 3,

where values of strictness levels are interpolated in-between any consecutive policy updates. Similarly, the outlier policies may form their own curves as denoted in the figure.

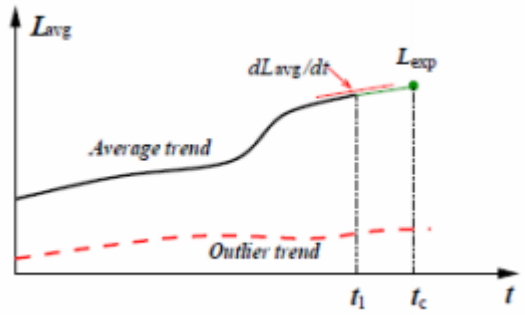


Fig. 3. Average Strictness Level Curve

A3P-SOCIAL

The A3P-social employs a multi-criteria inference mechanism that generates representative policies by leveraging key information related to the user's social context and his general attitude toward privacy. As mentioned earlier, A3Psocial will be invoked by the A3P-core in two scenarios. One is when the user is a newbie of a site, and does not have enough images stored for the A3P-core to infer meaningful and customized policies. The other is when the system notices significant changes of privacy trend in the user's social circle, which may be of interest for the user to possibly adjust his/her privacy settings accordingly. In what follows, we first present the types of social context considered by A3P-Social, and then present the policy recommendation process.

Modeling Social Context We observe that users with similar background tend to have similar privacy concerns, as seen in previous research studies and also confirmed by our collected data. This observation inspires us to develop a social context modeling algorithm that can capture the common social elements of users and identify communities formed by the users with similar privacy concerns. The identified communities who have a rich set of images can then serve as the base of subsequent policy recommendation. The social context modeling algorithm consists of two major steps. The first step is to identify and formalize potentially important factors that may be informative of one's privacy settings. The second step is to group users based on the identified factors. First, we model each user's social context as a list of attributes: {sc1, sc2, ..., scn}, where sci denote a social context attribute, and n is the total number of distinct attributes in the social networking site. These social context attributes are extracted from users' profiles. Besides basic elements in users' profiles, many social sites also allow users to group their contacts based on relationships (e.g., friends, family members). If such grouping functionality is available, we will consider its influence on privacy settings too. In a social site, some users may only have their family members as contacts, while some users may have contacts including different kinds of people that they met offline or on the Internet. The distribution of contacts may shed light on the user's behavior of privacy settings. We assume that users who mainly share images among family members may not want to disclose personal information publicly, while users having a large group of friends may be willing to share more images with a larger audience [19]. Formally, we model the ratio of each type of relationship among all contacts of a user as social connection. Let R1, ..., Rn denote the n types of relationships observed among all users. Let NuRi denote the number of user U's contacts belonging to relationship type Ri. The connection distribution (denoted as Conn) is represented as below:

$$Conn : \{\frac{N^u_{R_1}}{\sum_{i=1}^{n} N^u_{R_i}}, \ldots, \frac{N^u_{R_n}}{\sum_{i=1}^{n} N^u_{R_n}}\}.$$

For example, suppose that there are four types of relationships being used by users in the system: R1="family", R2="colleague", R3="friend", R4="others". Bob has 20 contacts, among which he has 10 family members, 5 colleagues, and 5 friends. His social connection is represented as { 10/20 , 5/ 20 , 5/ 20 , 0/ 20}. It is worth noting that, the number of social context attributes may grow when more rich information is collected by social networking sites in the future, and our algorithm is dynamic and capable of dealing with any number of attributes being considered. The second step is to identify groups of users who

have similar social context and privacy preference. Regarding social context, it rarely happens that users share the same values of all social context attributes. More common cases are that a group of users have common values for a subset of social context attributes. Such subset can be different for different groups of users, which makes the user grouping a challenging task. We illustrate the scenario using the following example. For simplicity of illustration, we take a smaller set of attributes to be considered. The obtained social groups have not taken into account privacy preferences yet. It is certainly possible users within the same social group maintain various privacy preferences. In order to tie social groups to privacy preferences, we further divide the social groups into sub-groups according to the closeness of their privacy preferences. In particular, we sort the users in the same social group in an ascending order of their privacy strictness levels. Then, starting from the user (say ui) with the minimum strictness level (Li), we scan the sorted list and include users whose strictness levels are no more than Li + ξ, where ξ is set to 4 the same as that in the policy prediction (in Section 4.2.2) for determining the closeness of the strictness levels. After one subgroup is formed, we remove the users in the subgroup from the sorted list. If the sorted list is not empty, we start the subgroup formation again in the same way until all users have been grouped.

Identifying Social Group We now introduce the policy recommendation process based on the social groups obtained from the previous step. Suppose that a user U uploaded a new image and the A3P-core invoked the A3P-social for policy recommendation. The A3P-social will find the social group which is most similar to user U and then choose the representative user in the social group along with his images to be sent to the A3P-Core policy prediction module to generate the recommended policy for user U. Given that the number of users in social network may be huge and that users may join a large number of social groups, it would be very time consuming to compare the new user's social context attributes against the frequent pattern of each social group. In order to speed up the group identification process and ensure reasonable response time, we leverage the inverted file structure [32] to organize the social group information. The inverted file maps keywords (values of social context attribute) occurring in the frequent patterns to the social groups that contain the keywords. Specifically, we first sort the keywords (except the social connection) in the frequent patterns in an alphabetical order. Each keyword is associated with a link list which stores social group ID and pointers to the detailed information of the social group. In the identified social group, we further examine its sub-groups by comparing the strictness levels of the subgroups with the new user's preferred privacy strictness level if provided. We select the sub-group whose strictness level matches the new user's privacy requirements best. If the new user did not specify privacy preference, we select the sub-group with the largest members. Then, in this selected sub-group, we look for the user who is most similar to the new user. We just need to compare the new user's and the group members' remaining attributes that are not included in the frequent pattern.

V.CONCLUSION

We have proposed an Adaptive Privacy Policy Prediction (A3P) system that helps users automate the privacy policy settings for their uploaded images. The A3P system provides a comprehensive framework to infer privacy preferences based on the information available for a given user. We also effectively tackled the issue of cold-start, leveraging social context information. Our experimental study proves that our A3P is a practical tool that offers significant improvements over current approaches to privacy.

REFERENCES

[1] A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In Privacy Enhancing Technologies Workshop, 2006.

[2] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In J.

B. Bocca, M. Jarke, and C. Zaniolo, editors, 20th International Conference on Very Large Data Bases, September 12-15, pages 487–499. Morgan Kaufmann, 1994.

[3] S. Ahern, D. Eckles, N. S. Good, S. King, M. Naaman, and R. Nair. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. In Conference on Human factors in computing systems, pages 357–366. ACM, 2007.

[4] M. Ames and M. Naaman. Why we tag: motivations for annotation in mobile and online media. In Conference on Human factors in computing systems, CHI' 07, pages 971– 980. ACM, 2007.

[5] A. Besmer and H. Lipford. Tagged photos: concerns, perceptions, and protections. In CHI '09: 27th international conference extended abstracts on Human factors in computing systems, pages 4585–4590. ACM, 2009.

[6] A. D. Bland JM. Multiple significance tests: the bonferroni method. BMJ, 310(6973), 1995.

[7] J. Bonneau, J. Anderson, and L. Church. Privacy suites: shared privacy for social networks. In Symposium on Usable Privacy and Security, 2009.

[8] J. Bonneau, J. Anderson, and G. Danezis. Prying data out of a social network. In ASONAM: International Conference on Advances in Social Network Analysis and Mining, pages 249– 254, 2009.

[9] O. Chapelle, P. Haffner, and V. Vapnik. Support vector machines for histogram-based image classification. Neural Networks, IEEE Transactions on, 10(5):1055–1064, 1999.

[10] H.-M. Chen, M.-H. Chang, P.-C. Chang, M.-C. Tien, W. H. Hsu, and J.-L. Wu. Sheepdog: group and tag recommendation for flickr photos by automatic search-based learning. In 16th ACM international conference on multimedia, pages 737–740. ACM, 2008.

[11] M. D. Choudhury, H. Sundaram, Y.-R. Lin, A. John, and D. D.Seligmann. Connecting content to community in social media via image content, user tags and user communication. In 2009 IEEE International Conference on Multimedia and Expo, ICME 2009, pages 1238–1241. IEEE, 2009.

Authors Profile:

K.Rajani M-Tech Dept. of CSE SreeVahini Institute of Science and Technology Tiruvuru Andhra Pradesh.

G.V.Ramana Associate.Professor M-Tech Dept. of CSE SreeVahini Institute of science and Technology Tiruvuru Andhra Pradesh.

J.V Krishna Associate.Professor M-Tech Dept. of CSE SreeVahini Institute of Science and Technology Tiruvuru Andhra Pradesh.