Enhanced privacy policy settings for user-uploaded images on content sharing sites

Ashwini N, Anmol Shyam, Hasan Matheen Baig, Tarun Kumar, Shreyas K S

ABSTRACT: With the increasing volume of images users share through social sites, maintaining privacy has become a major problem, as demonstrated by a recent wave of publicized incidents where users inadvertently shared personal information. In light of these incidents, the need of tools to help users control access to their shared content is apparent. 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. Over time, the generated policies will follow the evolution of users' privacy attitude. We provide the results of our extensive evaluation over 5,000 policies, which demonstrate the effectiveness of our system, with prediction accuracy over 90 percent.

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+, Flickr 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 Flickr 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. 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. 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. 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. However, existing proposals for automating privacy settings appear to be inadequate to address the unique privacy needs of images, due to the amount of information implicitly carried within images, and their relationship with the online environment wherein they are exposed. In this paper, we propose an Adaptive Privacy Policy Prediction (A3P) system which aims to provide users a hassle free privacy settings experience by automatically generating personalized policies. The A3P system handles user uploaded images, and factors in the following criteria that influence one's privacy settings of images:

II LITERATURE SURVEY

As in the paper [1] "Over exposed privacy patterns and considerations in online and mobile photo sharing", sharing personal media online becomes easier and widely spread, new privacy concerns emerge – especially when the persistent nature of the media and associated context reveals details about the physical and social context in which the media items were created. In a first-of-its-kind study, we use context-aware camera phone devices to examine privacy decisions in mobile and online photo sharing. Through data analysis on a corpus of privacy decisions and associated context data from a real-world system, we identify relationships between location of photo capture and photo privacy settings. Our data analysis leads to further questions which we investigate through a set of interviews with 15 users. The interviews reveal common themes in privacy considerations: security, social disclosure, identity and convenience. Finally, we highlight several implications and opportunities for design of media sharing applications, including using past privacy patterns to prevent oversights and errors.

As in the paper [2] "Privacy suites: shared privacy for social networks", creating privacy controls for social networks that are both expressive and usable is a major challenge. Lack of user understanding of privacy settings can lead to unwanted disclosure of private information and, in some cases, to material harm. We propose a new paradigm which allows users to easily choose suites" of privacy settings which have been specified by friends or trusted experts, only modifying them if they wish. Given that most users currently stick with their default, operator-chosen settings,

such a system could dramatically increase the privacy protection that most users experience with minimal time investment.

As in the paper [3] "Sheepdog: groups and Tag Recommendation for Photos by Automatic Search-based Learning", Online photo albums have been prevalent in recent years and have resulted in more and more applications developed to provide convenient functionalities for photo sharing. In this paper, we propose a system named Sheepdog to automatically add photos into appropriate groups and recommend suitable tags for users on Flickr. We adopt concept detection to predict relevant concepts of a photo and probe into the issue about training data collection for concept classification. From the perspective of gathering training data by web searching, we introduce two mechanisms and investigate their performances of concept detection. Based on some existing information from Flickr, a ranking-based method is applied not only to obtain reliable training data, but also to provide reasonable group/tag recommendations for input photos. We evaluate this system with a rich set of photos and the results demonstrate the effectiveness of our work.

As in the paper [4] "Personalizing Image Search Results on Flickr", The social media site Flickr allows users to upload their photos, annotate them with tags, submit them to groups, and also to form social networks by adding other users as contacts. Flickr offers multiple ways of browsing or searching it. One option is tag search, which returns all images tagged with a specific keyword. If the keyword is ambiguous, e.g., "beetle" could mean an insect or a car, tag search results will include many images that are not relevant to the sense the user had in mind when executing the query. We claim that users express their photography interests through the metadata they add in the form of contacts and image annotations. We show how to exploit this metadata to personalize search results for the user, thereby improving search performance. First, we show that we can significantly improve search precision by filtering tag search results by user's contacts or a larger social network that includes those contact's contacts. Secondly, we describe a probabilistic model that takes advantage of tag information to discover latent topics contained in the search results. The users' interests can similarly be described by the tags they used for annotating their images. The latent topics found by the model are then used to personalize search results by finding images on topics that are of interest to the user.

IV Current challenges:

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. Bonneau et al. 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 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. develop privacy settings based on a concept of "Social Circles" which consist of clusters of friends formed by partitioning users' friend lists. Ravichandrantal 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. Fangetal 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. 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 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, retrieval, and photo ranking, also in the context of online photo sharing sites, such as Flickr. Of these works, Zerr's work 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 versus public), so the classification task is very different than ours. Also, the authors do not deal with the issue of cold-start problem.

V EXISTING SYSTEM:

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. 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.

VI MOTIVATION

Sharing images within online content sharing sites, therefore may quickly lead to unwanted disclosure and privacy violations. 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. The aggregated information can result in unexpected exposure of one's social environment and lead to abuse of one's personal information.

VIIPROPOSED SYSTEM:

In this paper, we propose an enhanced privacy settings (A3P) system which aims to provide users a hassle free privacy settings experience by automatically generating personalized policies. The A3P system handles user uploaded images, and factors in the following criteria that influence one's privacy settings of images: The impact of social environment and personal characteristics. Social context of users, such as their profile information and relationships with others may provide useful information regarding users' privacy preferences. For example, users interested in photography may like to share their photos with other amateur photographers. The role of image's content and metadata. In general, similar images often incur similar privacy preferences, especially when people appear in the images. For example, one may upload several photos of his kids and specify that only his family members are allowed to see these photos.

Our work is related to some existing recommendation systems which employ machine learning techniques. Chen et al. proposed a system named Sheepdog to automatically insert photos into appropriate groups and recommend suitable tags for users on Flickr. They adopt concept detection to pre edit relevant concepts (tags) of a photo. Choudhury et al. proposed a recommendation framework to connect image content with communities in online social media. They characterize images through three types of features: visual features, user generated text tags, and social interaction, from which they recommend the most likely groups for a given image. Similarly, Yuetal. proposed an automated recommendation system for a user's images to suggest suitable photo-sharing groups. There is also a large body of work on the customization and personalization of tag-based information retrieval, which utilizes techniques such as association rule mining. For example, proposes an interesting experimental evaluation of several collaborative filtering algorithms to recommend groups for Flickr users. These approaches have a totally different goal to our approach as they focus on sharing rather than protecting the content.

Advantages of Proposed System:

The A3P-core focuses on analyzing each individual user's own images and metadata, while the A3P-Social offers a community perspective of privacy setting recommendations for a user's potential privacy improvement. We design the interaction flows between the two building blocks to balance the benefits from meeting personal characteristics and obtaining community advice.

VIII RESULTS

Users can view the images based on the policies that they are entitled to.

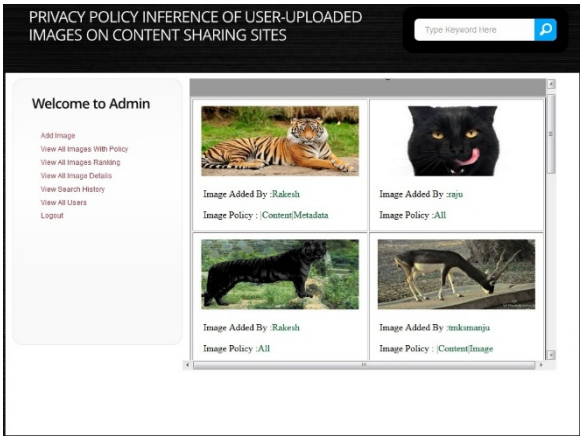

Figure:1 Policy of images.

The admin has privilege to view the search history of all the users, as shown in the figure below.

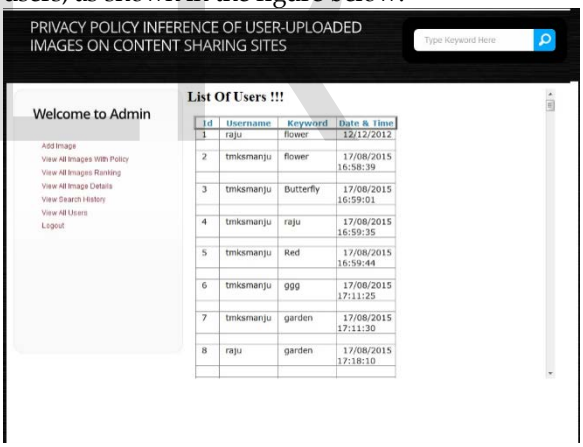

Figure:2 Search history.

Each image is given a rank based on the popularity of the image. This appears to the users when they view the image.

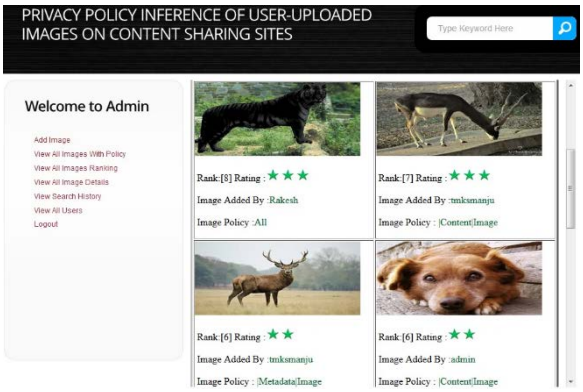

Figure: 3 Ranking of images.

When a user tries to upload an image, he may add some information along with those images (such as tags, content and metadata). He is also required to select the policy of the image.

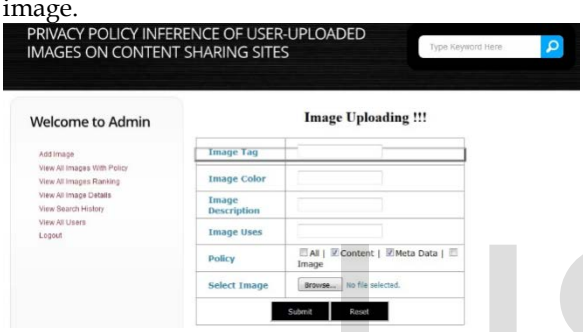

Figure: 4 Image uploading.

A user may search for images based on a few categories, content, tags and metadata.

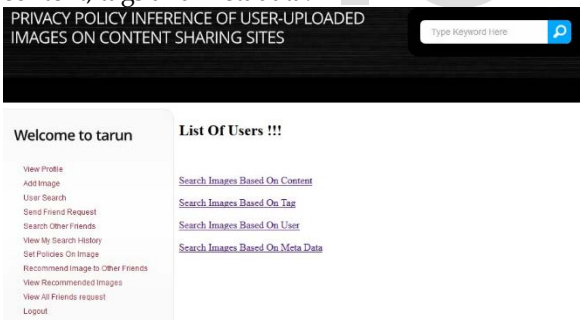

Figure: 5 Search options for users.

IX CONCLUSION

We have proposed an Enhanced privacy policy settings (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.

X REFERENCES

[1] A. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the facebook," in Proc. 6th Int. Conf. Privacy Enhancing Technol. Workshop, 2006, pp. 36–58.

[2] R. Agrawal and R. Srikant,"Fast algorithms for mining association rules in large databases," in Proc. 20th Int. Conf. Very Large Data Bases, 1994, pp. 487–499.

[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 Proc. Conf. Human Factors Comput. Syst., 2007, pp. 357–366.

[4] M. Ames and M. Naaman, "Why we tag: Motivations for annotation in mobile and online media," in Proc. Conf. Human Factors Comput. Syst., 2007, pp. 971–980.

[5] A. Besmer and H. Lipford, "Tagged photos: Concerns, perceptions, and protections," in Proc. 27th Int. Conf. Extended Abstracts Human Factors Comput. Syst., 2009, pp. 4585–4590.

[6] D. G. Altman and J. M. Bland ,"Multiple significance tests: The bonferroni method," Brit. Med. J., vol. 310, no. 6973, 1995.

[7] J. Bonneau, J. Anderson, and L. Church, "Privacy suites: Shared privacy for social networks," in Proc. Symp. Usable Privacy Security, 2009.

[8] J. Bonneau, J. Anderson, and G.Danezis, "Prying data out of a social network," in Proc. Int. Conf. Adv. Soc. Netw. Anal. Mining.,2009, pp.249–254.

[9] 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 Proc. 16thACM Int. Conf. Multimedia, 2008, pp. 737–740.

[10] 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 Proc. IEEE Int. Conf. Multimedia Expo, 2009, pp.1238–1241.

[11] L. Church, J. Anderson, J. Bonneau, and F. Stajano, "Privacy stories: Confidence on privacy behaviors through end user programming," in Proc. 5th Symp. Usable Privacy Security, 2009.

[12] R. da Silva Torres and A. Falc~ao, "Content-based image retrieval: Theory and applications," Revista de Inform_aticaTe_orica e Aplicada, vol. 2, no. 13, pp. 161–185, 2006.

[13] R. Datta, D. Joshi, J. Li, and J. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Comput. Surv., vol. 40, no. 2, p. 5, 2008.

[14] J. Deng, A. C. Berg, K. Li, and L. Fei-Fei, "What does classifying more than 10,000 image categories tell us?" in Proc. 11th Eur. Conf. Comput. Vis.: Part V, 2010, pp. 71–84. [Online]. Available: http://portal.acm.org/citation.cfm?id=1888150.1888157

[15] A. Kapadia, F. Adu-Oppong, C. K. Gardiner, and P. P. Tsang,"Social circles: Tackling privacy in social networks," in Proc. Symp. Usable Privacy Security, 2008.

[16] L. Geng and H. J. Hamilton, "Interestingness measures for data mining: A survey," ACM Comput. Surv., vol. 38, no. 3, p. 9, 2006.

[17] Image-net data set. [Online]. Available: www.image-net.org, Dec. 2013.

[18] S. Jones and E. O'Neill, "Contextual dynamics of group-based sharing decisions," in Proc. Conf. Human Factors Comput. Syst., 2011, pp. 1777–1786. [Online]. Available:http://doi.acm.org/10.1145/1978942.1979200

[19] A. Kaw and E. Kalu, Numerical Methods with Applications: Abridged., Raleigh, North Carolina, USA: Lulu.com, 2010. [20] 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. ACM Annu. Conf. Human Factors Comput. Syst., 2012, pp. 377–386.

[21] K. Lerman, A. Plangprasopchok, and C. Wong, "Personalizing image search results on flickr," CoRR, vol. abs/0704.1676, 2007.

[22] H. Lipford, A. Besmer, and J. Watson, "Understanding privacy settings in facebook with an audience view," in Proc. Conf. Usability, Psychol., Security, 2008.

[23] D. Liu, X.-S. Hua, M. Wang, and H.-J. Zhang, "Retagging social images based on visual and semantic consistency," in Proc. 19thACM Int. Conf. World Wide Web, 2010, pp.1149–1150.

[24] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Analyzing facebook privacy settings: User expectations vs. reality," in Proc. ACMSIGCOMMConf. Internet Meas. Conf., 2011, pp. 61–70.

[25] D. G. Lowe, (2004, Nov.). Distinctive image features from scale-invariant key points. Int. J. Comput. Vis. [Online]. 60(2), pp. 91–110. Available: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94

[26] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 8, pp. 959–973, Aug. 2003.

[27] E. M. Maximilien, T. Grandison, T. Sun, D. Richardson, S. Guo, and K. Liu, "Privacy-as-a-service: Models, algorithms, and results on the Facebook platform," in Proc. Web 2.0 Security Privacy Workshop, 2009.

[28] A. Mazzia, K. LeFevre, and A. E.,, "The PViz comprehension tool for social network privacy settings," in Proc. Symp. Usable Privacy Security, 2012.

[29] M. Rabbath, P. Sandhaus, and S. Boll, "Analysingfacebook features to support event detection for photo-based facebookapplications," in Proc. 2nd ACM Int. Conf. Multimedia Retrieval, 2012, pp. 11:1–11:8.

[30] R. Ravichandran, M. Benisch, P. Kelley, and N. Sadeh, "Capturing social networking privacy preferences," in Proc. Symp. Usable Privacy Security, 2009.

[31] A. Singhal, "Modern information retrieval: A brief overview," IEEE Data Eng. Bullet., Speicial Issue on Text Databases, vol. 24, no. 4, pp. 35-43, Dec. 2001.

[32] A. C. Squicciarini, S. Sundareswaran, D. Lin, and J. Wede, "A3p: Adaptive policy prediction for shared images over popular content sharing sites," in Proc. 22nd ACM Conf. Hypertext Hypermedia, 2011, pp.261–270.

- *Ashwini N is currently working as Assistant Professor, Department of ISE, BMS Institute of Technology & Mgmt., Bengaluru, Karnataka, E-mail: ashwinilaxman@bmsit.in*
- *Anmol Shyam is currently pursuing Under graduate degree program in Department of ISE, BMS Institute of Technology & Mgmt., Bengaluru, Karnataka , E-mail: anmols.shyam@gmail.com*
- *Hasan Matheen Baig is currently pursuing Under graduate degree program in Department of ISE, BMS Institute of Technology & Mgmt., Bengaluru, Karnataka , E-mail: hasanbaig31@gmail.com*
- *Tarun Kumar is currently pursuing Under graduate degree program in Department of ISE, BMS Institute of Technology & Mgmt., Bengaluru, Karnataka , E-mail: kumar.tarun097@gmail.com*
- *Shreyas K S is currently pursuing Under graduate degree program in Department of ISE, BMS Institute of Technology & Mgmt., Bengaluru, Karnataka , E-mail: shreyas.kasetty@gmail.com*