An A3P approach towards Image Privacy policy recommendation on content sharing sites

Mrs. Swapna R, Dr. Nagesh H R

[1] M.Tech Student, 2 Professor ,Dept. Of CSE , Mangalore Institute of Technology and Engineering, Mangalore, Karnataka, India

***Abstract*— Popularity of the social media like Facebook, Flickr, LinkedIn and others has increasing everyday due its various features provided. Some of the features are to collaborate with friends, good and friendly user interface to share data and multimedia content at ease. Data uploaded to Social network by the user are at high risk of vulnerability. User uploaded data on the social media does play vital role in user connectivity. Thus maintaining privacy of the user uploaded data on content distribution network is at most important. To control privacy of the uploaded data user should also provide privacy information for the uploaded data. It is observed the difficulty in setting privacy for images shared by user in social network. This requires automation privacy recommendation for the user shared images based on the image metadata. This suggested privacy inference for uploaded image should suits his satisfaction level. The necessary aspect considered are social context and image metadata for analyzing the privacy setting for the user uploaded data. Two-level framework is proposed which utilizes user's history for recommending best policy setting for uploaded image. Solution provided in paper relies on an categorizing the image based on image metadata which may belong to similar policies and predicting the policy using apriori data mining algorithm to automatically suggest a policy for each image uploaded by the user in social network. The goal of this paper is to provide extensive method for privacy policy suggestions to improve the security of shared data in the social media sites.**

***Keywords*— Content sharing sites, Apriori data mining algorithm, Social media, privacy inference.**

I. INTRODUCTION

The term ―Social network sites‖ refers to web-based services which individuals are allowed to create a user profile, and connects with list of users with whom user likes to share information, and view information of the other individuals who are already in the social network sites. Social media consumes lot of time, but still internet users are happy to use social networking sites. This will lead to the use of applications such as Facebook and Twitter becoming regular browsers, forming regular tendency that influence their daily activities in lives, both online and offline. Facebook, YouTube, Google+, Twitter and LinkedIn collectively have billions of users, and they're getting bigger and more important each and every year. Users more often utilizes online social network for establish communication with other users. Online social networks are websites which allow users to establish connections and build relationships to other Internet users. Social networking is used to be in contact with friends by building communication, make new contacts in social network of other users and find people with similar interests and ideas. Some of the tasks performed by the user are to share any special events, update user current location, share any invitation messages, current status update, and express their hobbies, wishing other person on occasion of any event and others. Most of the above task makes use of text, images or videos for sharing through the social network. The uploaded user information in social networks is stored remotely.

Information uploaded to the Social networks, stores data remotely on the server located in some other location, compare to user's personal computer. The relationship between privacy and a person's data in content sharing sites is variable. Thus it requires providing more security mechanism for the data uploaded by the user in content sharing sites. Privacy is essential to the design of security mechanisms to avoid unnecessary disclosure and violation of the privacy. Most of the content sharing site provider has offered settings to privacy. Those privacy settings can deny or allow others to access information of individual users. Individual users share their user data to the list of users depending on the content of the uploaded data. If the data uploaded to sites contains for example hobby related than data uploaded by the user to network can be shared to all in the site. If uploaded data is regarding any particular event than it will be shared with the specific group it is involved. Most content sharing sites allow individual users to set their privacy preferences. Unfortunately, some of the recent surveys have revealed that users struggle to organize and manage such privacy settings [9]. Hence, many have concluded the need of policy setting recommendation systems which can be convenient for users to configure privacy settings at ease. [2], [4], [10]. However, existing systems for automating privacy settings recommendation appear to be failing to address the individual privacy needs of images [5] considering the amount of data carried within the images, and their relationship with the online environment wherein they are exposed.

By considering required goal, we present an Adaptive Privacy Policy Prediction which focus to maintain users a settings recommendation by generating personalized policies automatically. Some of the important factors that have been considered to handle user uploaded images in A3P that influences privacy settings are:

- The privacy policy setting recommendation of user images uploaded can be maintained based on the user social activities and individual user personal characteristics.
- Metadata of the images helps in suggesting privacy settings.

II. RELATED WORK

Anna CinziaSquicciarini developed an Adaptive Privacy Policy Prediction (A3P) [1] system, a automated privacy settings recommendation for system generating personalized policies. The A3P system handles user uploaded images based on the person‘s personal characteristics and images metadata. The A3P system consists of two components: A3P Core and A3P Social. When a user uploads an image, the image will be first sent to the A3P-core. The A3P Core uses metadata for classification of the images and uses association rule data mining algorithm for police prediction if sufficient use data available in the system. Else if not sufficient data, A3P system will be called to fetch relevant policies. The disadvantage is inaccurate privacy policy generation in case of the absence of metadata information about the images. Also manual creation of metadata log data information leads to inaccurate classification and also violation privacy.

Jonathan Anderson proposed a tool called **Privacy Suites** [2]. This privacy suit allows users to easily choose ―suites" of privacy settings. A privacy suite can be created by an expert using privacy programming. Privacy Suites could also be created directly through existing configuration UIs or exporting them to the abstract format. The privacy suite is distributed through existing distribution channels to the members of the social sites. The disadvantage of a rich programming language is less understandability for end users. Given a sufficiently high-level language and good coding practice, motivated users should be able to verify a Privacy Suite. The main goal is transparency, which is essential for convincing influential users that it is safe to use.

Alessandra Mazzia introduced PViz Comprehension Tool [5], an interface and system that corresponds more directly with how users model groups and privacy policies applied to their networks. PViz allows the user to understand the visibility of her profile according to automatically-constructed, natural sub-groupings of friends, and at different levels of granularity. Because the user must be able to identify and distinguish automatically-constructed groups, we also address the important sub-problem of producing effective group labels. PViz is better than other current policy comprehension tools Facebook's Audience View and Custom Settings page.

Chen, Chang Proposed a system named SheepDog [3] to automatically insert photos into appropriate groups and recommend suitable tags for users on Flickr. They adopt concept detection to predict relevant concepts (tags) of a photo. A reliable system toadd photos into popular groups automatically. The system recommends suitable tags for photos and provides a user friendly interface such that users could easily select their favorite tags to attach

III. METHODOLY

The proposed A3P system consists of two main components: A3P-core and A3P-social. The overall data flow as shown in Fig 1 is described as follows. 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.

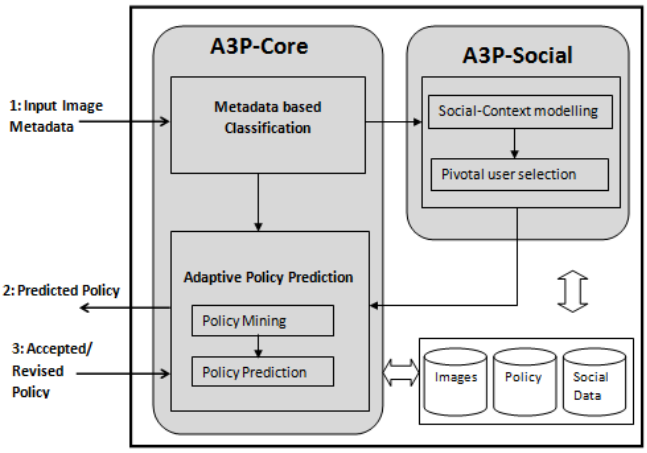

Fig.1. System Overview

A. A3P Core

The A3P Core consists of two major blocks of the framework.

- Metadata based Image Classification
- Adaptive Policy Prediction

Every image of the user gets classified based on the metadata and then its privacy policies are generalized. With the help of this approach, the policy recommendation becomes easy and more accurate. Based on the Classification based on metadata the policies are applied to the right class of images. Moreover combining the image and classification and policy prediction would enhance the system‘s dependency.

1) Metadata based Image Classification

Metadata can be fetched for the uploaded images in content sharing network. In general, similar images often incur similar privacy preferences.

2) Adaptive Policy Prediction

This section deals with the privacy concerns of the user by deriving the privacy policies for the images. The Adaptive Policy Prediction consists of two following sub-parts

- Policy Mining
- Policy Prediction

Policy mining deals with data mining of policies for similar categorized images and Policy prediction applies prediction algorithm to predict the policies.

- *Policy Mining*

The privacy policies are the privacy preferences expressed by the users. Policy mining deals with mining of these policies by applying different Apriori mining rules and steps. It follows the order in which a user defines a policy and decides what rights must be given to the images. This hierarchical mining approach starts by looking the popular subjects and

their popular actions in the policies and finally for conditions. It can be thoroughly reviewed with the help of following steps.

Step 1 of this process applies Apriori based data mining on the subject components of the policies of the new image. With Apriori based data mining we select the best frequent itemset.

Step 2 of this process applies a priori based data mining on the action components. Similar to the first step we will select the best frequent itemset which will give most popular combinations of action in policies.

Step 3 of this process mine the condition component in each policy set using a priori based data mining technique. The best frequent itemset are selected which gives us a set of attributes which often appear in policies.

Apriori algorithm is easy to execute and very simple, is used to mine all frequent itemsets in database. The algorithm makes many searches in database to find frequent itemsets where k-itemsets are used to generate k+1-itemsets. Each k-itemset must be greater than or equal to minimum support threshold to be frequency. Otherwise, it is called candidate itemsets. In the first, the algorithm scan database to find frequency of 1-itemsets that contains only one item by counting each item in database. The frequency of 1-itemsets is used to find the itemsets in 2-itemsets which in turn is used to find 3-itemsets and so on until there are not any more k-itemsets.

Some of the drawbacks of Association rule algorithm is that for each candidate itemset, there are as many entries as its support value that results in unnecessarily generating and counting too many candidate itemsets that turn out to be small. Apriori is used through reducing the time consumed in transactions scanning for candidate itemsets by reducing the number of transactions to be scanned. The Apriori algorithm takes advantage of the fact that any subset of a frequent itemset is also a frequent itemset. The algorithm can therefore, reduce the number of candidates being considered by only exploring the itemsets whose support count is greater than the minimum support count. All infrequent itemsets can be pruned if it has an infrequent subset. Apriori algorithm are more efficient and less time consuming.

- *Policy Prediction*

The policy mining phase may give us many policies but system needs to show the best one to the user. Thus, this approach is used to choose the best policy for the user by obtaining the strictness level. The Strictness level decides how ―strict‖ a policy is by returning an integer value. This value should be minimum to attain high strictness. The strictness can be discovered by two metrics: a major level and coverage rate. The major level is determined with the help of combinations of subject and action in a policy and coverage rate is determined using the condition statement. Different integer values are assigned according to the strictness to the combinations and if the data has multiple combinations we will select the lowest one. Coverage rate provides a fine-grained strictness level which adjusts the obtained major level. For example a user has to 5 friends and two of them are females. Hence if he specifies policy as ―friends‖=male, then the coverage rate can be calculated as (3/5) =0.6. Hence, the image is less restricted if the coverage rate value is high.

B. 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 intwo scenarios. One is when the user is a newbie of a site. The other is when the system notices significant changes of privacy trend in the user‘s social circle. In what follows, first presents the types of social context considered by A3P-Social, and then present the policy recommendation process.

1) Modeling Social Context

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.

2) Identifying Social Group

The policy recommendation process is based on the social groups obtained from the previous step. Suppose User U uploaded a new image. 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

C. Image metadata

User gives Image metadata as input to the A3p System. This image metadata will be stored in the database for further analysis.

D. Policy

After analyzing the policy from policy prediction using image metadata, policy will further be stored in database. This helps to retrieve this policy for future analysis of the policy for other metadata uploaded by the user.

DI. Social Data

Social Data is referred to user personal information and social relationship of the user in the social network sites. This is data is used in A3P Social method to form user communities based on social context attribute and Social connection in social networking sites.

V. CONCLUSION

The proposed system for suggesting privacy setting for uploaded images helps user to set privacy at ease. In A3P, the user uploads an image which passes through a classification engine. The engine retrieves similar images based on their

content. Images that have tags falling in the same group are considered similar. Adaptive policy setting works on the premise that users set similar policies for similar photos. Once the similar images are obtained, we use a policy mining algorithm based on A priori rule mining to predict the most likely policy, which is presented to the user for approval, and modification, if necessary. The A3P system helps to mine the policies based on user available history to infer privacy policy for uploaded images. Even the social context attributes are considered during policy evaluation in case of user history unavailability.

ACKNOWLEDGEMENT

I am very thankful to my guide Dr. Nagesh H R Head of the Department, Department of Computer Science and Engineering, Mite for his cordial support, valuable information and constructive suggestions during the planning and development of this work.

REFERENCES

[1] Anna CinziaSquicciarini, Member, IEEE, Dan Lin, SmithaSundareswaran, and Joshua Wede, ―Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites‖, IEEE Transactions on Knowledge and Data Engineering, Vol. 27, NO. 1, January 2015.

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

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

[4] 4. KambizGhazinour, Stan Matwin and Marina Sokolova, ―Yourprivacyprotector: A Recommender System For Privacy Settings In Social Networks‖, International Journal of Security, Privacy and Trust Management (IJSPTM) Vol 2, No 4, August 2013.

[5] 5. Alessandra Mazzia Kristen LeFevre and Eytan Adar, The PViz Comprehension Tool for Social Network Privacy Settings, Tech. rep., University of Michigan, 2011.

[6] 6. Peter F. Klemperer, Yuan Liang, Michelle L. Mazurek, ―Tag, You Can See It! Using Tags for Access Control in Photo Sharing‖, Conference on Human Factors in Computing Systems, May 2012.

[7] 7. C. A. Yeung, L. Kagal, N. Gibbins, and N. Shadbolt, ―Providing access control to online photo albums based on tags and linked data,‖ in Proc. Soc. Semantic Web: Where Web 2.0 Meets Web 3.0 at the AAAI Symp., 2009, pp.

[8] 8. SergejZerr, Stefan Siersdorfer, Jonathon Hare, Elena Demidova , I Know What You Did Last Summer!:Privacy-Aware Image Classification and Search , Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, 2012.

[9] 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.

[10] 13. K. Strater and H. Lipford, ―Strategies and struggles with privacy in an online social networking community,‖ in Proc. Brit. Comput.