International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2494-9150 Vol-02, Issue 03, June 2016.

Data Security Policies Inference on Content Sharing Sites

[1]Prachiti S. Pimple, [2]Prof. B. R. Nandwalkar

[1]M. E. Student, [2]Assistant Professor, [1,2]Late G.N.SapkalCOE, Nashik, Maharashtra, India.

[1]*prachiti.pimple@gmail.com*, [2]*nandwalkar.bhushan@gmail.com*

Abstract : Nowadays Social media has become extremely popular. We can communicate with a huge number of people using it. The social networking sites such as LinkedIn and Facebook, have given individuals the opportunity to meet new people and make friends across the world. People who use these social networking platforms share a lot of personal information with a large number of "friends." Situation of people sharing personal images on these type of sites create a serious privacy concern, Which needs to be taken care of in order to improve the satisfaction level of a particular user. Hence, we have decided to develop a system that will help the user to maintain security for images he/she has uploaded on a content sharing site.

Keywords — ***Content sharing sites, Privacy, Social media***

I. INTRODUCTION

Users are able to build connections and relations to other Internet users using these online social networks. The information stored on these sites is not stored on the personal computer of a respective user, but on a distant space. One can use such social networking sites to connect to people who share common interests and share similar ideas.

It is required to deploy security mechanisms for the development of online social networks. Privacy is also an integral part in designing the security mechanisms. Most of the social network providers offer privacy settings in order to permit or completely deny the access to personal details of a particular user. In certain conditions we wish to share some information amongst ourselves only inside a small circle of close friends and not to anyone else. But there are other instances, when we willingly reveal private information to anonymous outsiders, but not to the people who know us better or vice versa. Let us discuss about Internet Privacy now. To put easily, it can be said to be the ability of user to control [1] what information one reveals about oneself, and [2] who can access that information. When mentioned simply, privacy is violated, when the data is analyzed or collected without the owner knowing about it. The purposes or the reason for which data is being or will be used should be made available to the user. A user is allowed to enter his/her privacy preferences on most sites, whose primary objective is to share content. But the studies show that many users are unaware of the fact that they can manage such preferences and hence the system in turn fails. This process is not known to be fool proof and that is the main concern. Hence the policy [10], [11]. Therefore, many have turned to the need of policy recommended systems that support a user to easy and hassle free configure privacy settings [2],[3],[4]. However, existing proposed system for automating privacy settings have turned out to be insufficient to solve the unique problem of privacy needs of images, [5] because to the amount of information that is exclusively carried within images, and their relationship with the online environment wherein they are exposed.

There are two strategies for giving privacy to the user information. 1. User can himself enter his privacy preferences 2. Using of recommended systems settings that assist users to set his privacy preferences. Users' social environment and personal characteristics determine the privacy policy of the data uploaded by the user. Social data that a user updates i.e. their profile information and relationship status with others users may provide useful information about a privacy preferences of a particular user. Content of an image and metadata provide privacy policy of an image that is uploaded by a user. A hierarchical image classification which classifies

images first based on their contents and then refines each category into subgroups based on their metadata. Images that do not contain any metadata are usually grouped by content only. Such a hierarchical classification provides a higher priority to image content and minimizes the influence of missing tags.

II. LITERATURE SURVEY

A. *Adaptive Privacy Policy Prediction*

Adaptive Privacy Policy Prediction (A3P) [1] system, a free privacy settings system automatically generate personalized policies. The A3P system handles user uploaded images based on the person's personal characteristics and images content and metadata The disadvantage is inaccurate privacy policy generation in case of the absence of metadata information about the images.

B. *Privacy Suites*

Privacy Suites [7] allows users to easily choose "suites" of privacy settings. A privacy suite is collection of different privacy policies. Privacy suite can be created by professional created using privacy programming. Privacy Suites could also be created directly through existing configuration UIs or exporting them to the abstract format The drawback of a rich programming language is less understandability for end users.

C. *Social Circle*

Social circles [8] provide a web based solution to protect personal information. The technique named Social Circles Finder automatically generates the friend's list. It is a technique that analyses the social circle of a person and identifies the intensity of relationship and therefore social circles provide a meaningful categorization of friends for setting privacy policies.

D. *Privacy Protector*

Recommender system YourPrivacy Protector [2] understands the social net behaviour of their privacy settings and recommending reasonable privacy options. It uses user's personal profile, User's interests and User's privacy settings on photo albums as parameters and with the help of these parameters the system constructs the personal profile of the user. It automatically learned for a given profile of users and assigns the privacy choices.

E. *PViz Comprehension Tool*

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.

In existing system users struggle to set up and maintain privacy settings. 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 proposed system we are implementing policy recommendation systems which can assist users to easily and properly configure privacy settings.

III. PROPOSED SYSTEM

All the Methods like privacy suites, social circle that are discussed in the literature survey have some drawbacks. The A3P system i.e. Adaptive Privacy Policy Prediction system helps the users to compose privacy settings for their images uploaded on social networking sites.A3P system 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.A3P system consist of two component 1)A3P core 2) A3P social .when user upload an image ,it will sent to A3P core.A3P core classifies the images based on their content like size, texture and metadata like tags ,comments. Here for the extraction of the features of images we are using SURF algorithm. A3P core will decide whether to invoke A3P social. 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. In revise policy user can define new policy for uploaded image. He/she can set different permission like view, tag, and comment, like, share or combination of different permission to the different user.

Algorithm:

Input:

Image

Output:

Predicted policy

Begin

1) Content Based Image Classification is performed

2) Metadata based classification is performed

3) Identification of social groups

4) Adaptive policy prediction

End

A. *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. 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.

1) Content based classification:

When a user uploads an image, it is handled as an input query image. The newly uploaded image is compared with the 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.

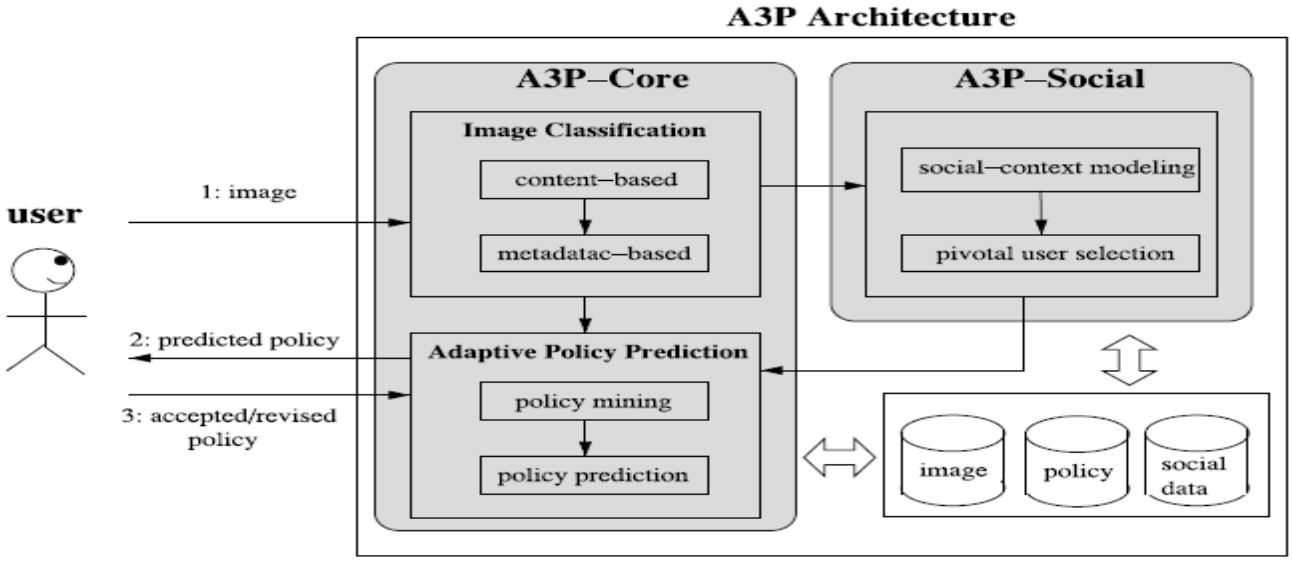

Fig 1. A3P Architechture

Speeded Up Robust Features (SURF)

For features extraction of images we are using SURF algorithm.

a] Find image interest points by using Hessian matrix

b] Find major interest points in scale space by expressing Hessian as Taylor expression and calculate $\frac{\partial h}{\partial x}$

c] Find feature direction by using Haar wavelet transform.

d] Generate feature vectors.

2) Metadata based classification:

Steps:

1) Extract keywords from the metadata associated with an image.

2) Identify all the nouns, verbs and adjectives in the metadata and store them as metadata vectors tnoun = {t1;t2;...;ti}, tverb = {t1;t2;...; tj} and tadj ={t1;t2;...;tk}

3) Derive a representative hypernym

i. Select the hypernym

ii. there are more than one hypernyms with the same frequency, consider the hypernym closest to the most relevant baseline class to be the representative hypernym.

4) Find a subcategory that an image belongs to

i. first image forms a subcategory as itself and the representative hypernyms of the image becomes the subcategory's representative hypernyms.

ii. Compute the distance

$$Dist_{m=w_n}.D(h_n, h_n^c) + w_a.D(h_a, h_a^c) + w_v.D(h_v, h_v^c)$$

5) Check for the closest subcategory based on distance value. The new image will be included in to the subcategory

6) Update the representative hypernyms of the subcategory by keeping the hypernyms with the highest frequency.

Otherwise, a new subcategory will be constructed for this image.

Support Vector Machine (SVM):

SVM is used for training purpose.

CandidateSV ={closest pair from opposite classes}

While there are violating points do
Find a violator
CandidateSV = candidateSV U violator
if any αp < 0 due to addition of c to S then
CandidateSV = candidateSV \p
Repeat till all such points are pruned
end if
end while

B.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.

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.

IV . RESULTS

Here we give the image as an input and output for this is number of policies. For every uploaded image by user numbers of policies are predicted. When we test our system we got the accuracy of correctly predicting a policy is 90% where the accuracy we got while testing using existing approach is 70%.

Existing	Proposed
70	90

Table no. 1. Accuracy obtained by existing approach and proposed approach in percentage

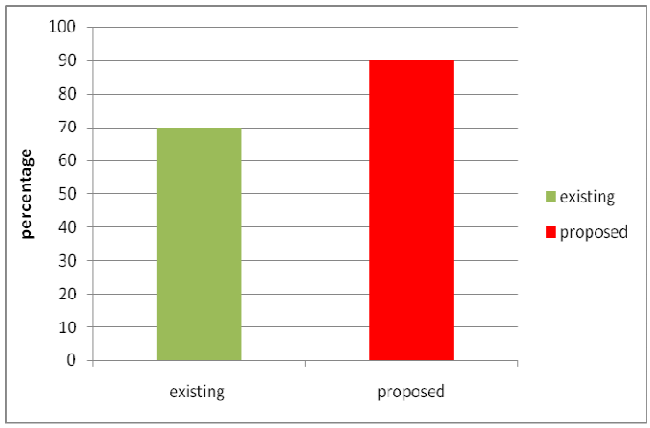

Fig 2. Accuracy obtained by existing approach and proposed approach in percentage

Content Based	Metadata based	Content with Metadata
70	90	90
80	90	95

Table no.2.Accuracy obtained by different method in percentage

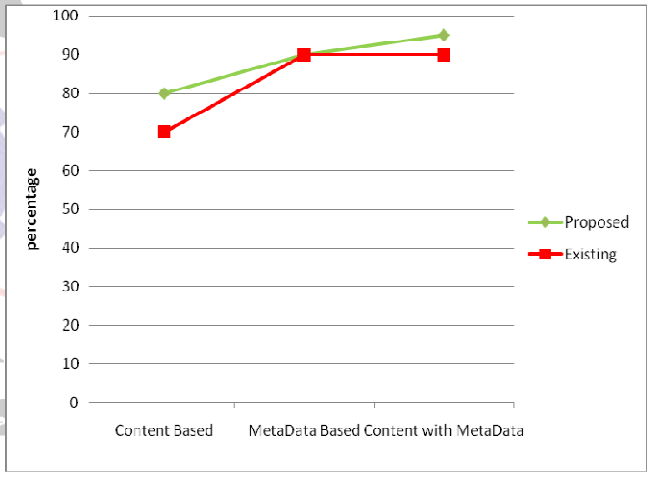

Fig 3. Accuracy obtained by different method in percentage

From the above table we can see the difference generated for the results of different approaches for proposed and existing system. Table No.1 shows the accuracy obtained by existing approach as well as proposed approach .Table No.2 shows accuracy obtained by individual approaches .In existing system when we used content based approach ,metadata based approach and combination of content and metadata based approach accuracy obtained is like 70%,90%,90%. With same approach in proposed system accuracy obtained is like 80%, 90%, 95%.

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 framework to infer privacy preferences based on the information available for a given user. We improve the accuracy for prediction of privacy policy for image that are uploaded by content sharing site user. If user uploads an image and he wants to set privacy for that then he can allow the particular user which is present in his friend list to tag, share, download, like permission. It predicts accurate results as compare to existing system. Our experimental study proves that proposed system offers significant improvements over current approaches to privacy.

ACKNOWLEDGMENT

With all respect and gratitude, I would like to thank all people who have helped us directly or indirectly for the research. I would like to thank my guide, Prof.B.R.Nandwalkar, for his guidance and support. I will forever remain grateful for the constant support and guidance extended by guide, in making this project. Through our many Discussions, he helped me to form and solidify ideas. The invaluable discussions I had with him, the penetrating questions he has put to me and the constant motivation, has all led to the development of this research.

REFERENCES

[1] Anna Cinzia Squicciarini, Member, IEEE, DanLin, Smitha Sundareswaran, 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] Kambiz Ghazinour, Stan Matwin And Marina Sokolova, "Yourprivacy protector: A Recommender System For Privacy Settings In Social Networks", International Journal of Security, Privacy and Trust Management (IJSPTM) Vol 2, No 4, August 2013.

[3] Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, Elena Demidova, I Know What You Did Lass summer!: Privacy -Aware Images Classification and Search , Proceedings of the 35th International ACM SIGIR conference on Research and development in information retrieval, 2012.

[4] 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.

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

[6] C. A. Yeung, L. Kagal, N. Gibbins, and N. Shadbolt, "Providing Access control to online photo albums base on tags and Linked data," in Proc. Soc Semantic Web: Where Web 2.0Meets Web 3.0 at the AAA Symp. 2009.

[7] J. Bonneau, J. Anderson, and L. Church, "Privacy suites: Shared Privacy for social Networks, " in Proc. Symp. Usable Privacy Security, 2009 .J . Breckling, Ed., The Analysis *of* Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics.Berlin, Germany: Springer, vol.61.

[8] 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.

[9] 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.

[10] H. Lipford, A. Besmer, and J. Watson, "Understanding privacy Settings in Facebook with an audience view," Proc. Conf. Usability, Psychol., Security, ss2008

[11] A. Acquisti and R.Gross, "Imagined Communitie Awareness, information sharing, and privacy on the facebook, in Proc. 6th Int .Conf.

