Privacy Policy Inference for Social Images

Chaitrali Salunke
ME (Computer Engg.)
VPCOE, Baramati,
Savitribai Phule University
Pune, Maharashtra

S. A.Takale
Asst. Prof. IT Dept.
VPCOE, Baramati,
Savitribai Phule University
Pune, Maharashtra

ABSTRACT

Social networking has become an indivisible part of the modern lifestyle. It plays a crucial role in our daily lives. It allows us to communicate with numerous individuals.Popular social networking sites such as Facebook1, LinkedIn2 allow people to get connected from diverse geographic locations. Existing and the emerging social media sites empower users to utilize various interesting features such as sharing photos or information with as many friends as they want, commenting on a text or picture, creating groups and so on. Privacy settings offered by these sites come into picture when a user may not want to share his profile globally or with certain people. Privacy violation is an important issue that needs to be addressed while being active on any social networking site. If the privacy settings provided by the respective site are inadequate, then people may invade your privacy and misuse your information. Going towards this need, utilizing system to compose privacy settings for user's images is important. Further, the privacy inference policies should be maintained with respect to user profile. 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

Hierarchical clustering, Cosine Similarity, Social media, Content sharing, Privacy Policies, User Profiles.

1. INTRODUCTION

Creating privacy controls for social media or networks that are both expressive and usable is a major challenge. The word "social media" refers to the group of Internet-based and mobile services that allow users to participate in online exchanges, bring user-created content, or join online communities. Online social networks or sharing/content sites are websites that allow users to build or construct connections and relationships to other Internet users. Social networks store information remotely, rather than on a user's personal computer. Social networking can be used to keep in touch with friends, make new friends and find people with similar interests and ideas. The relation between privacy and a person's social network is multi-faceted. So it required to develop more security mechanisms for different communication technologies, especially online social networks. Privacy is very important to the design of security mechanisms. Most social networks providers have provide an opportunity of privacy settings to allow or refuse others access to personal information details. In certain event or an occasion we want information about ourselves to be known only by a small circle of close friends, and not by strangers or unknown people. In other side, we are willing to reveal our personal information to strangers, but not to those who know us better. Social network theorists have studied the relevance of relations of different depth and strength in a person's social network and the valued of so-called weak ties in the flow of information across different nodes in a network. Internet privacy can be defined as the ability to control what information one reveals about oneself, and who can access that information. Essentially, when the data is gathered or analyzed without the knowledge or permission of its owner, privacy is violated.

When it comes to the usage of the data, the owner should be informed about the purposes and aim for which the data is being or will be used. Most of the Content sharing or Photo sharing websites permit users to choose their privacy preferences.

Unfortunately, recent studies have shown that users struggle or it is difficult 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 tiresome and error-prone. Therefore, many have recognized the need of policy recommendation systems which can help users to easily and properly configure privacy settings. However, existing system for automating privacy settings appear to be insufficient to address the unique privacy needs of images due to the amount of information absolutely carried within images and their relationship with the online environment wherein they are exposed. The privacy of user data can be given by using three methods or approaches.

1. The user can set the privacy preferences
2. Usage of recommendation systems which helps users for setting the privacy preferences.
3. Users can block unwanted contents or can report abuse to site owner.

The privacy policies of user uploaded data can be provided on the user social environment and personal characteristics. Social context of users, such as their profile information and relationships with others may provide beneficial information regarding users' privacy preferences. The privacy policy for image which is uploaded by user can be provided by the user uploaded image content and its metadata. A hierarchical image classification which classifies images based on their contents and then decides each category into subcategories based on their metadata. Images that do not have metadata will be classified together only by content. Hierarchical classification provides by A3P system which gives a higher priority to image content and reduces the influence of missing tags [1].

2. LITERATURE SURVEY

Many studies and analysis have been performed on privacy policy techniques. Alessandra Mazzia et al. [3] introduced PViz Comprehension Tool, an interface and system that relates all the more straightforwardly corresponds more directly with how users model groups and privacy policies applied to their networks. Also this tool permits the user to understand the Visibility of her profile according to automatically-constructed groupings of friends.

Peter F. Klemperer et al. [4] developed a tag based access control of data shared in the online networking locales. A system that creates access-control policies from photo management tags. Every photo is consolidated with an access grid for mapping the photo with the participant's friends. The participants can select a suitable preference and access the information. Photo tags can be categorized as organizational or communicative based on the user needs.

Sergej Zerr et al. [5] proposed a technique Privacy-Aware Image Classification and to enable privacy-oriented image search. It combines textual Meta-data images with variety of visual features to provide security policies. Search to naturally recognize private pictures, and to empower security situated picture seek. It consolidates literary meta information pictures with assortment of visual components to give security approaches.

Choudhury et al. [6] proposed a recommendation framework to connect image content with groups in online social networking. They characterize images through three sorts of features: visual features, user generated text tags, and social interaction, from which they recommend the most likely groups for a given image.

Jonathan Anderson et al. [7], proposed a paradigm called Privacy Suites which permits users to effortlessly pick "suites" of security settings. A security suite can be made by a specialist utilizing protection programming. Privacy Suites could also be created directly through existing configuration UIs.

Danezis et al. [8] proposed a machine-learning based way to deal with consequently separate security settings from the social connection inside of which the information is delivered. It creates privacy settings taking into account an idea of "Social Circles" which consist of clusters of friends formed by partitioning users' friend lists.

Fabeah Adu-Oppong et al. [9]. It Created security settings in light of the idea of social circles. It gives an electronic answer for secure individual data. The procedure named Social Circles Finder, consequently creates the companion's rundown. It is a strategy that investigations the social circle of a man and recognizes the power of relationship and thusly social circles give a significant arrangement of companions for setting security approaches.

Kambiz Ghazinour et al. [10] composed a recommender framework known as Your Privacy Protector that comprehends the social net behavior of their privacy settings and recommending reasonable privacy options. It uses user's personal profile, User's interests and User's privacy settings on photograph collections as parameters and with the assistance of these parameters the system constructs the personal profile of the user.

Fong et al. [11] proposed a privacy wizard to help client's award benefits to their companions. 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.

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.

3. PROBLEM STATEMENT

Let U= $\{u_0,u_1\ldots u_{n-1}\}$ be the set of users, A=$\{a_0,a_1..a_i\}$ and I= $\{i_0,i_1\ldots i_n\}$ be the set of actions of user and the set of shared images respectively. Then the system requires to calculate set P= $\{p_0,p_1..p_{n-1}\}$ which contents the privacy policies maintained by the system. Further probability P_i can be defined as

$$P_i = \begin{cases} true, & if\ isMatch(image, profiles, action) \\ & false, \quad otherwise \end{cases}$$

The system should generate an alert message if privacy predicts a non matching contents which then user may allow or disallow.

4. MOTIVATION

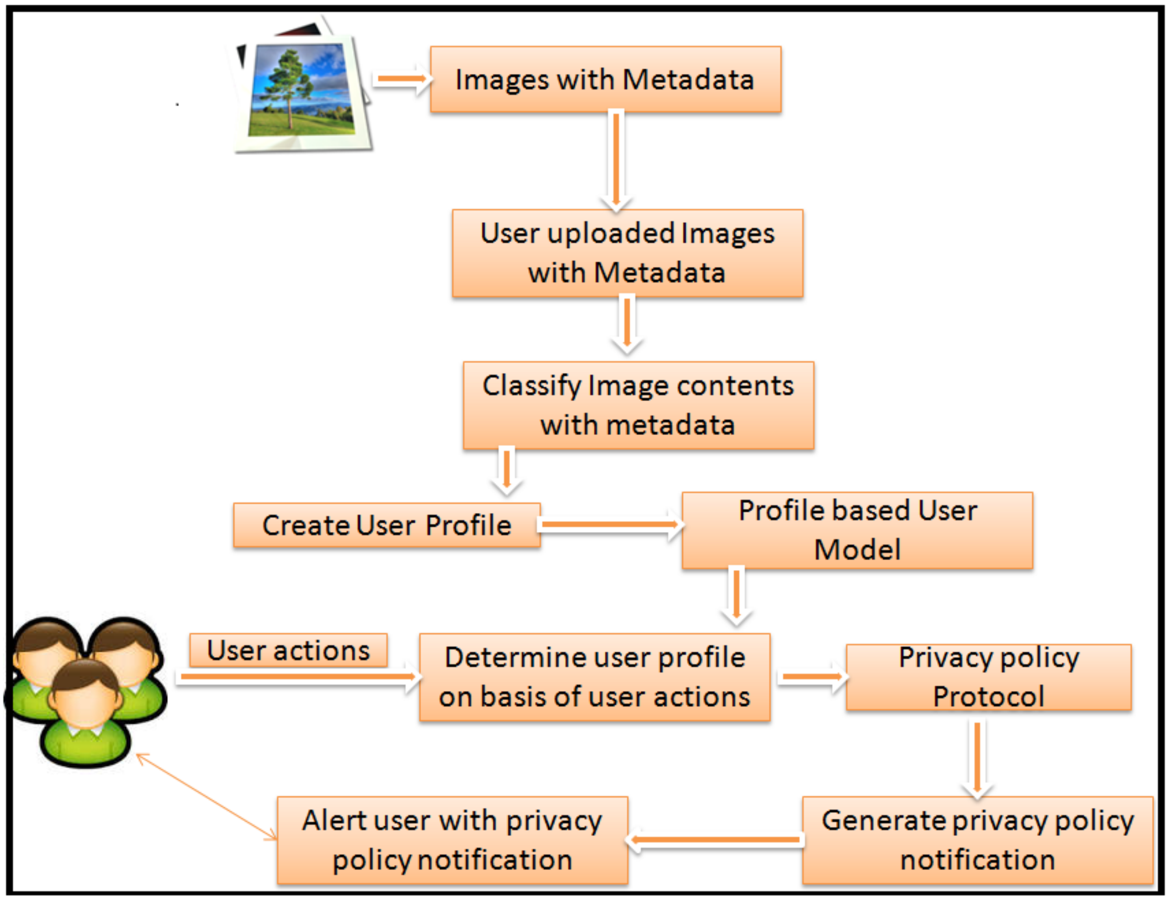
As earlier system have some disadvantages like less understandability of methods like privacy suits due to high privacy programming etc. So here solution depends on image classification framework for image categorize which are having similar policies. System generate user profile model and maintained privacy policies.

5. PROPOSED METHOD

We propose a user profile model which aims to provide users privacy setting experience by providing policies. Also to provide more security to images uploaded by user compare to other system. In general, similar images often incur similar privacy preferences, especially when people appear in the images. Using User profile model, Photo sharing/ content sharing websites allow or maintain privacy for the User Profile instead for only contents which leads more security. That means when user upload an image ,it will sent to our System. The System 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 Single hierarchical algorithm.

The User actions i.e. (view, comment, download) are either public/private, while system maintain privacy policies for each action. Fig.1 shows system architecture. For each user, his/her images are first divided on the basis of content and metadata. Then, privacy policies of each category of images are analyzed for the policy prediction. Using 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. The two-stage approach allows the system to employ the first stage to classify the new image and find the candidate sets of images for the subsequent policy recommendation.

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. Our main contribution to the existing work is to generate user profile, further the privacy inference policies should be maintained with respect to user profile.


Images with Metadata
User uploaded Images with Metadata
Classify Image contents with metadata
Create User Profile
Profile based User Model
User actions
Determine user profile on basis of user actions
Privacy policy Protocol
Generate privacy policy notification
Alert user with privacy policy notification

6. IMPLEMENTATION DETAILS

This section gives idea about the strategy of implementation. In this system user group socially connected to each other and also shared some contents like images etc. Suppose user upload images and metadata then our system goes into some steps which describes below.

The system can be divided into five stages.

1. Image Classification
2. Profile Generation
3. Profile based user model
4. Privacy Policy
5. Alert and Generate Notification

6.1 Image Classification

In this stage, after uploading process system processes the metadata. This stage classifies or to obtain groups of images that may be associated with similar privacy preferences, System propose a hierarchical image classification i.e hierarchical clustering which classifies images first based on their contents and then divide each category into subcategories based on their metadata.

So here image classifies on the basis of

1) Content-based classification
2) Metadata-based classification

6.1.1 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. Content-based classification is based on an efficient and accurate image similarity approach. For this classification color histogram is used. A color histogram[1] is a representation of the distribution of colors in an image. For digital images, a color histogram represents the number of pixels that have colors in each of a fixed list of color ranges that span the image's color space, the set of all possible colors. The color histogram can be built for any kind of color space, although the term is more often used for three-dimensional spaces like RGB or HSV. In general, a color histogram is based on a certain color space, such as RGB or HSV. When we compute the pixels of different colors in an image, if the color space is large, then we can first divide the color space into certain numbers of small intervals. Each of the intervals is called a bin. This process is called color quantization. Then, by counting the number of pixels in each of the bins, we get the color histogram of the image.

In brief, first extract color values from an image. For this we segmented one image into many small pieces and then for each and every piece the red, green and blue color values are extracted. After this process we took an average of red, green and blue values for one small image. So after this process we get three color values (red, green and blue) for each small part of our main storable image. This color values are also stored in three 1D arrays. we have to convert our image to a grey scale one. Though the color values of the image is lost but that makes our edge detection more efficient. After conversion and edge detection is done we have taken the maximum edge value for one column of the image and stored it into a 1D array. Finally after all the feature values are extracted and stored for one image we get five 1D arrays for one image.the pixels of different colors in an image, if the color space is large, then we can first divide the color space into certain numbers of small intervals. Each of the intervals is called a bin. This process is called color quantization. Then, by counting the number of pixels in each of the bins, we get the color histogram of the image.

6.1.2 Metadata-Based Classification

The metadata-based classification groups images into subcategories under aforementioned baseline categories.

1] The first step is to extract keywords from the metadata associated with an image.

2] The second step is to generate tokens from it. Tokens (t_i).

$$\text{COSINE}(\text{Text}_1 ,\text{Text}_2)= \frac{\sum(Titext1)*(Titext2)}{\sqrt{(Titext1)2}\sqrt{(Titext2)2}}$$

For matching tokens the regular expression is used. After classification of images i.e. clusters of images are generated on the basis of contents and metadata. These clusters are updated by the time of uploading process. Hierarchical clustering is used for this image clustering which describes below:

Steps:

```
1. Create a new cluster $C_0$ .  (where $C_0$ is primary cluster)
2. Add $I_0$ To $C_0$ . (where $I=\{I_0,I_1,..I_r\}$ is set of images)
3. for each remaining $I_r$ do
    flag=0;
    mostmatching = -1;
      for each $C_j$  in C do
       calculate sim($C_j$ , $I_r$ )  (where $C=\{C_0,C_1,\ldots C_j,..C_n\}$)
         if (sim $\geq$ thr && sim $\geq$ flag) then
            flag= sim;
            mostmatching= j;
     if mostmaching $\neq$-1 then
     Add $I_r$ To $C_j$
     else
    Create a new cluster $C_{n.}$
    Add $I_r$ To $C_n$
4. Stop.
```

6.2 Profile Generation

As after collection of images and metadata the preprocessing is done and it generates tokens. In the classification different classes are generated on the basis of similar policies. By the time classes are updated whenever users upload various images on sites. On the basis of classification images and metadata system generates the user profiles having similar policies of images which are uploaded by user.

In this metadata of images that is all the strings collected in a list which is uploaded by one user and stored in clusters form. After taking union of tokens, which stored in new list Occurrence of tokens in list is calculated and using the Turn Frequency (1) that is TF formula the name of user profile is generated.

$$\text{TF}(t_i) = \frac{\#\,occurences\ of\ ti}{\#tokens} \quad \_\_\_\_\_(1)$$

6.3 Profile based user model

In this stage, After generation of user profiles, system generates user notification wall for each user who socially connected to each other. In this first step is to load user profiles from database. The actions (tag, comment, view download and upload) determine user profile, so for each action profile list is generated like uploaded Profile, view Profile, download Profile etc. contents and metadata the user profile is Generated on the basis of user actions i.e. tag, comment, download and view. Classifier generates the classes using cosine similarity. On the basis of classification images and metadata system generates the user profiles having similar policies of images which are uploaded by user. Using this Data, profile based model is created.

So after classification of images that is clusters are generated on the basis of contents and metadata. These clusters are updated by the time of uploading process. After classification user profiles are generated. Hierrachical clustering_1 is used for this image clustering.

6.4 Privacy Policy

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.

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. 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.

After the user profile generation process and according to the user actions i.e. Tag, view, comments etc. the privacy policy is applied to same category of new image. In simple way the mining is to follow natural order in which user defines policy. Given an image, a user usually first decides who can access the image, and then thinks about what specific access rights (e.g., view only or download) should be given.

6.5 Alert and Generate Notification

In this stage, 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. That means a user usually first decides who can access the image, then the image, then thinks about what specific access rights(e.g. view , comment etc.) should be given, and finally refine the access conditions. System generates notification whether he/ she has permission or not on that particular actions.

7. RESULTS

We evaluate the effectiveness of our Proposed system in terms of the policy prediction accuracy and user acceptability. The result reveals that in the earlier system privacy policies are applied to contents which are uploaded by user. But in this, system generates privacy policies to each user profile instead of contents which provides more security.

For System the input data is shown in Table 1. i.e. number of images or contents shared by multiple users.

Table 1: Input Dataset

#users	#content shared	#classes
10	30	5
20	70	7
30	50	7

Result evaluation for our System is shown in Table 2, number of users who uploads different kind of images or shared their images. Number of actions performed on it i.e. view, comment, tag, upload and download.

Table 2: Result Evaluation

User Actions	#TCA	#UC	#CNM	#PP	PR	RC
View	20	5	4	3	0.80	0.75
Comment	32	22	17	16	0.77	0.94
Upload	50	12	9	8	0.75	0.88
Tag	30	14	12	10	0.85	0.83
Download	22	20	18	15	0.90	0.83

#TCA= #total content shared #UC= #user contents #PP= #privacy prediction #CNM=#content with no match PR=Precision, RC=Recall

Each user was given a distinct set of images taken from the Picalert project data set including Flickr images on various subjects and different privacy sensitivity. On average, each user labeled images with their policies and added two to three tags each.

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%.

Table no. 3. Accuracy obtained by existing approach and proposed approach in percentage

Existing	Proposed
70	90

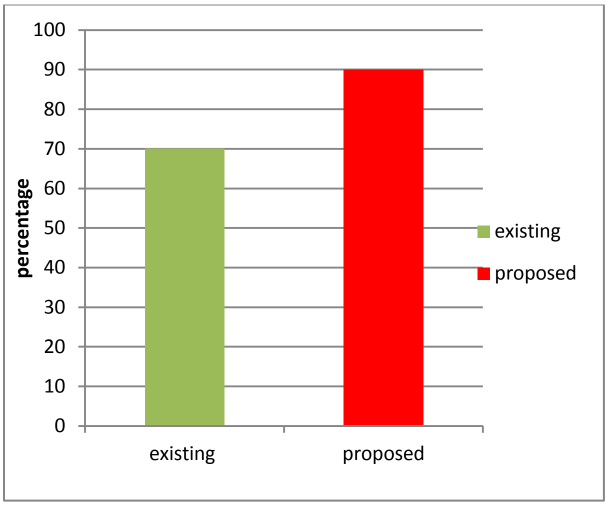


Fig 2. Accuracy obtained by existing approach and proposed approach in percentage

From the above table no.3 we can see the difference generated for the results of different approaches for proposed and existing system. Table No.3 shows the accuracy obtained by existing approach as well as proposed approach .Table No.4 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%.

Table no.4 Accuracy obtained by different method in percentage

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

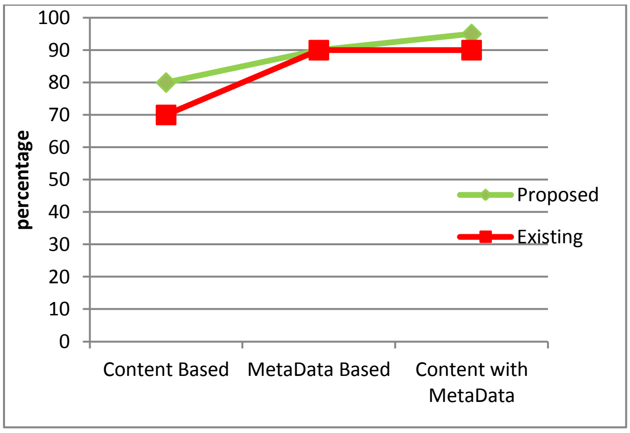


Fig 3. Accuracy obtained by different method in percentage

7.1 Precision and Recall

Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance In this system precision and recall is calculated as follows:

$$Precision = \frac{No\ of\ image\ contents\ with\ no\ match}{No\ of\ images\ shared}$$

$$Recall = \frac{No\ of\ correct\ predictive\ images}{No\ of\ images\ shared}$$

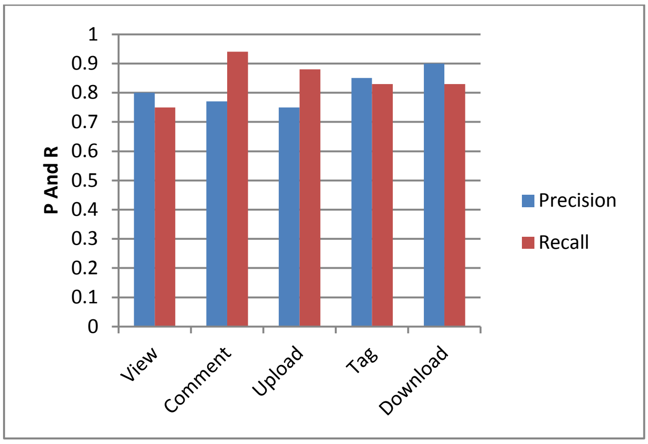


Fig 4. Graph of precision and Recall

The Fig 4 shows graph of precision and recall of each action of user that is view, comment, upload, tag and download.

8. CONCLUSION

In this paper, we propose a profile based user model to help users to integrate privacy settings for shared contents in Content Sharing Websites. For photo sharing and content sharing sites it helps to maintain or allow privacy policies for user profile instead of their contents. Conventional content sharing websites like Flickr[1] and Facebook[2] contributes the privacy policies using either public/private privacy settings, while the proposed system maintains privacy policies using user profiles and user actions. Our main contribution to the existing work is to generate user profile using Hierarchical Clustering, further the privacy inference policies are maintained with respect to user profile. It predicts accurate results as compare to existing system. Our experimental study proves that proposed system offers significant improvements over current approaches to privacy.

9. ACKNOWLEDGMENTS

I would like to express my deepest appreciation to all those who provided their valuable guidance to me. A special gratitude I give to my guide Prof. S.A.Takale for her contribution in stimulating suggestions and encouragement. Furthermore I would also like to acknowledge with much appreciation the crucial role of the Head of Department and Principal of our college. Also, thank to the anonymous reviewers for their valuable comments.

10. REFERENCES

[1] Anna Cinzia Squicciarini, Member, IEEE, Dan Lin, Smitha Sundareswaran, and Joshua Wede, "Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites" IEEE Transaction On Knowledge And Data Enginnering, VOL. 27, NO. 1, January 2015 193

[2] Miss.Chaitrali A. Salunke Student, Computer Engineering Department, V.P College of Engineering, Baramati "A Survey On Privacy Policy Inference for Social Images" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 03 Issue: 02 | Feb-2016

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

[4] 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.

[5] S. Zerr, S. Siersdorfer, J. Hare, and E. Demidova, "Privacy-aware image classification and search," in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2012, pp. 35–44.

[6] H. Sundaram, L. Xie, M. De Choudhury, Y. Lin, and A. Natsev, "Multimedia semantics: Interactions between content and community," Proc. IEEE, vol. 100, no. 9, pp. 2737–2758, Sep. 2012.

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

[8] 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. 9–14.

[9] 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.

[10] K. Ghazinour, M. Sokolova, and S. Matwin. "Detecting Health-Related Privacy Leaks in Social Networks Using Text Mining Tools." Advances in Artificial Intelligence. Springer Berlin Heidelberg, 2013. 25-39

[11] P. Fong, M. Anwar, and Z. Zhao. A privacy preservation model for facebook-style social network systems. University of Calgary Technical Report 2009-926-05, 2009

[12] R. da Silva Torres and A. Falc~ao, "Content-based image retrieval: Theory and applications," Revista de Inform_atica Te_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.