ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 12, December 2018

Enhanced Privacy Policy Prediction for User-Uploaded Images on Content Sharing Sites

Rupali Narkhede[1], Madhuri Zawar[2]

P.G. Student, Department of Computer Engineering, G's COE, Jalgaon , North Maharashtra University, Maharashtra India[1]

Professor, Department of Computer Engineering, G's COE, Jalgaon , North Maharashtra University, Mahashtra India [2]

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 accuracies over 90 percent.

KEYWORDS: Adaptive privacy policy prediction , content sharing sites, metadata, online information services, web based services**.**

I. INTRODUCTION

The internet and online social networks, in particular, are a part of most people's lives. Emarketer.com reports that in 2011, nearly 150 million us internet users will interface with at least one social networking site per month. Emarketer.com also reports that in 2011, 90 percent of internet users ages 18-24 and 82 percent of internet users ages 25-34 will interact with at least one social networking site per month. This trend is increasing for all age groups. As the young population ages, they will continue to leverage social media in their daily lives. In addition, new generations will come to adopt the internet and online social networks. These technologies have become and will continue to be a vital component of our social fabric, which we depend on to communicate, interact, and socialize.

Not only are there a tremendous amount of users online, there is also a tremendous amount of user profile data and content online. For example, on Facebook, there are over 30 billion pieces of content shared each month. New content is being added every day; an average Facebook user generates over 90 pieces of content each month. This large amount of content coupled with the significant number of users online makes maintaining appropriate levels of privacy very challenging.

In addition, it measures the human effects of our improvements. It introduces three new improvements to privacy management models:

1. Assisted Friend Grouping—an incremental improvement to traditional group-based policy management.
2. Same-As Policy Management—a new paradigm improvement over traditional group-based policy management.
3. Example Friend Selection—an incremental improvement to Same-As Policy Management.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 12, December 2018

The report leverages traditional group-based policy management as our baseline and progressively improve upon this privacy management model. With each new enhancement, we measure their human effects including cluster/user defined relationship group alignment, user privacy sentiment, efficiencies and user perceptions. The report introduces a user-assisted friend grouping mechanism that enhances traditional group-based policy management approaches. Assisted Friend Grouping leverages proven clustering techniques to aid users in grouping their friends more effectively and efficiently. It introduces a new privacy management model that is an improvement over traditional group-based policy management approaches. The new paradigm leverages a user's memory and opinion of their friends to set policies for other similar friends, which we refer to as Same-As Policy Management. Users associate the policy with an example friend and in doing so have this friend in the forefront of their mind. This allows users to be more selective and careful in assigning permissions. Users are thinking of people, not groups. Using a visual policy editor that takes advantage of friend recognition and minimal task interruptions, Same-As Policy Management demonstrated improved performance and user perceptions over traditional group-based policy management approaches.

II. RELATED WORK

This section outlines the social media ecosystem and provides an overview of the privacy problems faced by social media users. It starts with section 2.1.1 where the various stakeholders of a typical social media ecosystem are described before providing a sys-tematization of literature in section 2.1.2 which helps in enumerating the previous work done in the area of social media privacy. This helps in identifying the gaps in previous work and positioning the work presented in this thesis in the broader spectrum of social media privacy.

In addition to systematizing existing relevant literature, this section also looks at the access control and contact grouping mechanisms available in the popular social media sites to enhance the understanding of state of the art access control mechanisms available to social media users. Section 2.2 provides an evaluation of 30 popular social media sites in terms of the support provided to users for defining and maintaining social relationships by the means of the access control and contact grouping mechanisms afforded to them by the social media infrastructure.

A. Social Media Ecosystem and Stakeholders

Before detailing the various privacy problems faced by social media users, it is important to understand the ecosystem of a typical social media site. Most popular social media sites have three types of primary stakeholders:

- Users create profiles on the social media sites by providing their information such as personal details (name, age, location, etc.), photos, multimedia content, etc. They are also able to share information such as text, photos, etc., in the form of ``posts'''' or ``updates'''' with people who they connect with on these sites. Most of the content on social media sites is created by users which makes them a key element of the social media ecosystem.
- Providers run the social media infrastructure, store users' information and manage its distribution to other users and third parties. Notably, providers are in charge of enforcing their own privacy policies while also communicating these policies to the users in the form of documentation as well as privacy controls. Most popular social media sites have a centralized structure where the provider is in charge of the entire infrastructure. There are some decentralized social networks such as Diaspora[1]and Friendica[2]but are not as extensively used when compared to popular centralized social networks such as Facebook and Google+.
- Third parties are neither users nor providers. They add to the basic functionality of social media sites by providing services to the users. For example, many popular social media sites allow third parties to provide applications such as social games which can be used by the users to add to their overall social experience.

In terms of the overall ecosystem of social media sites, users need to interact with both the social media Providers as well as the Third-parties via the user interface of the social media site. They use these interfaces to update their profile, connect with people by adding them as ``friends'''' (or ``connections'''', ``followers'''', etc.) and interact with these friends by sharing content with them and making access control decisions. The access control behavior of the users depends to a large extent on the mechanisms offered to them by the social media site. This is discussed in more detail in Section 2.2.

B. Privacy Problems in Social Media

The social media activity of users often entails disclosure of personal information and hence brings with it risks and threats to their privacy. Such social media privacy problems can be, and indeed have been, categorized in many different ways. Such categorizations and classifications help in an overall understanding of the overall spectrum of privacy problems in social media. For this chapter, the social media privacy problems are classified in two categories:

- Social Privacy problems arise due to the interaction between social media users on the social media sites. The most important aspect of social media for users is the ability to interact with vast networks of friends who represent different life facets (such as friends, family, co-workers, etc.). In such a situation, failure to appropriately control access to their content can lead to information being revealed to unintended audiences which may lead to a breach of their privacy. Such situations may arise either due to the users' failure to configure access controls appropriately or even in a situation where they made the appropriate access control decisions but actions by the members of their audience result in a breach of privacy.
- Institutional Privacy problems arise due to the social media infrastructure provided to users. These can be due to the business interests of the social media providers or simply failure to address conceptual gaps in the infrastructure or policy which leaves the privacy of the users at a risk. For this classification, institutional privacy will encompass threats arising directly due to the social media infrastructure as well as the third party applications used by social media users. A distinction is not made between these two for this classification as they often overlap, for example, privacy policies created by social media providers govern how user data can be used by third parties. Moreover, the users often fail to distinguish between third parties and the social media infrastructure and have demonstrated disapproval of third-party access to their data [MC10] while also freely using such applications on their social media profiles[3].

C. Institutional Privacy Problems

The privacy of social media users may be put at risk due to the practices and policies of the social media providers and these problems, classified as institutional privacy problems, are discussed here.

Policy Deficiency : The policies set out by the social media provider, which govern how user data can be collected, fail to safeguard users' privacy.

Description: Privacy policies provided to social media users may be considered deficient and insufficient as users may simply avoid reading what they consider as legalese and even when they do read them, may fail to understand the contents [SSM11, FFB15]. Moreover, privacy policies are often incomplete as they acknowledge nominal mechanisms behind data collection by a provider or a third party, but do not sufficiently elaborate the privacy implications necessary for users to make informed decisions [S$^+$10]. Another problem is that privacy policies may not be correctly implemented at the infrastructure level as this can be a challenging engineering task [AHB04] due to the evolving nature of policies which are prone to modifications over time.

Solutions: Possible enhancements of privacy policies include creating machine-readable privacy policies aimed to find a match with users' privacy settings [Cra03] and better representation of the users' privacy preferences using a ``privacy persona'''' [SHC$^+$09].

Access Control in Current Social Media Infrastructures

In addition to understanding the various privacy threats and mitigation detailed in previous work mentioned in existing literature, it is also essential to examine the status-quo in terms of the access controls afforded to users by the current social media infrastructures in order to provide a holistic analysis. This is especially important as the nature of such mechanisms often shape the users' ability to safeguard their privacy.

This section provides a systematic evaluation of the top 30 social media sites [Ale14] (as of January, 2015) and classifies them into categories offering similar support to users by modeling social contexts through contact grouping mechanisms. Dating sites, online shopping sites and sites that were too specific to particular populations (for example, Classmates for US graduates and Naijapals for Nigerians) were excluded from this evaluation.

III. Proposed Algorithm

A. Flow of Image Uploading System
1. START
2. Select an image to upload by the login user.
3. Enter appropriate title for the selected image.
4. Process to upload the image in the system.
5. Call method to get image id which is having most similar heading and suitable names.
(Algorithm of Privacy Policy)
6. Get privacy policies already set for the result image unique identity.
7. Shows policies to user.
8. If user is satisfied with policies then continue to upload image.
9. If user is not satisfied with policies then allow user to set privacy policy for the image and continue to upload.
10. STOP

IV. PSEUDO CODE

A.*Algorithm of Privacy Policy Prediction*
INPUT: Caption & Tags.
OUTPUT: Relevant Image Id.
1. Get headings and names from front-end.
2. Execute SQL query to search for image having exact same caption and tags.

```
Resultset matchId=executeQuery(ExactMatch(WholeHeading&& AllNames));
If(matchId is not null)
{
    Return matchedId;
}Else{
    Resultset matchId=executeQuery(ExactMatch(WholeHeading|| AllNames));
    If(matchId is not null){
        Return matchId;
    }else{
        Return 0;
    }
}
```

V.SIMULATION RESULTS

Performance Evaluation
For Performance evaluation of the approach we measure it based on 2 parameters i.e precision and recall . Precision and Recall are defined in terms of a set of retrieved documents (e.g. the list of documents produced for a query) and a set of relevant documents (e.g. the list of all documents that are relevant for a certain topic)

Precision :

Precision is the fraction of retrieved documents that are relevant to the find.

$$\text{Precision} = \frac{|\{\text{Relevant document}\}\cap\{\text{retrived document}\}|}{|\{\text{retrived document}\}|}$$

Recall :

Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$\text{Recall}=\frac{|\{\text{Relevant document}\}\cap\{\text{retrived document}\}|}{|\{\text{relevent document}\}|}$$

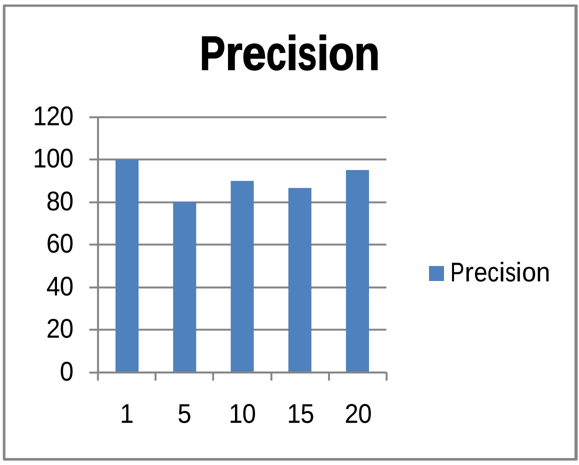
F-Measure:

Measure is the information retrieval measure from precision and recall.

$$\text{F-Measure}=\frac{2*(\text{precision}*\text{recall})}{(\text{precision}+\text{recall})}$$

Total Relevant	Retrieved	Relevant Retrieved	Precision	Recall	F-Measure
1	1	1	100	100	100
5	5	4	80	80	80
10	10	9	90	90	90
15	15	13	86.6667	86.7	86.66666667
20	20	19	95	95	95

Table 1. Result of direct user evaluation.



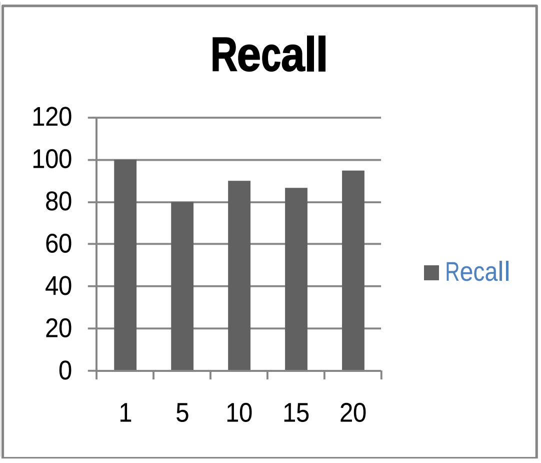


Fig 1. Coloum chart according to precision value

Fig 2. Coloum chart according to recall value

DOI: 10.15680/IJIRCCE.2018. 0612017

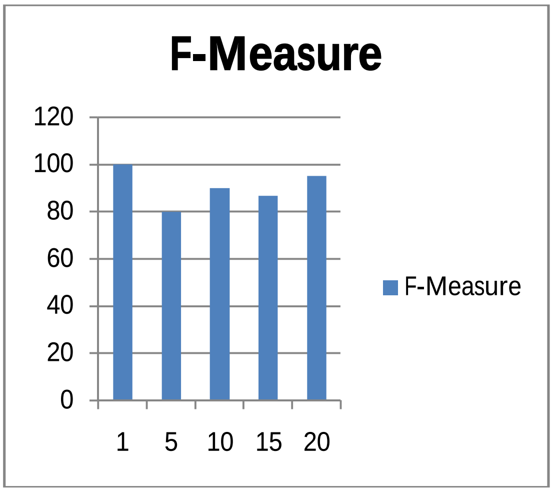


Fig 3. Coloum chart according to F-Measure value

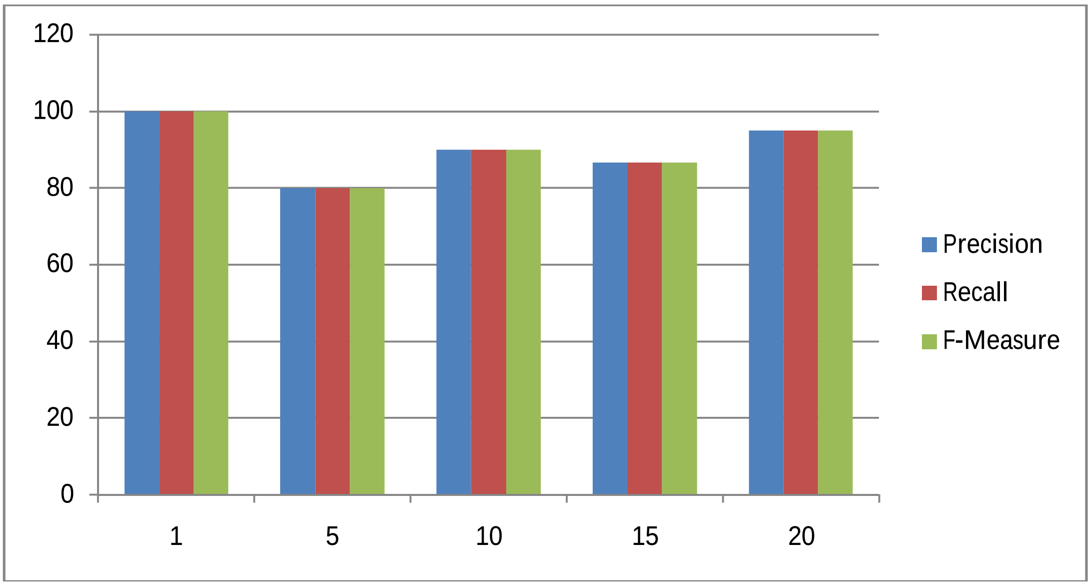


Fig 4. Comparative performance.

VI. CONCLUSION AND FUTURE WORK

We have proposed Policy Prediction system that helps users automate the privacy policy settings for their uploaded images and videos. The 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.

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. 16th ACM 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_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.

[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.