A Frame Work for Preserving Privacy in Social Media using Generalized Gaussian Mixture Model

P Anuradha
Department of CSE
GITAM University
Visakhapatnam, India

Y.Srinivas
Department of IT
GITAM University
Visakhapatnam, India

MHM Krishna Prasad
Department of CSE
J NT University,
Kakinada

***Abstract*—Social networking sites helps in developing virtual communities for people to share their thoughts, interest activities or to increase their horizon of camaraderie. Social networking sites come under few of the most frequently browsed categories websites in the world. Nevertheless Social Networking sites are also vulnerable to various problems, threats and attacks such as disclosure of information, identity thefts etc. Privacy practice in social networking sites often come into sight, as information sharing stands in conflict with the disclosure-related misuse. Face book is one such most popular and widely used Social Networking sites which have its own robust set of Privacy mechanisms. Yet they are also prone to various privacy issues and attacks. The impulse in this paper lies in proposing a novel approach for improving the privacy among the social networking sites .The article presents the issues by a novel approach based on tagging and a model based technique based on generalized Gaussian Mixture Model.**

***Index Terms*—Privacy; Social Network; Social Relevant Groups; Generalized GMM, Tagging**

I. INTRODUCTION

The recent advancements in technologies, in particular, the internet technology, benefitted the users to share and view useful information from across the globe. Thousands of clients share their thoughts online through the social network sites[1],[2],[3]. With the mammoth options available on Social Networking sites, it creates a Virtual world for the users. Social Networking sites go upwards because of all these reasons. These technologies help to share the information by the individuals and experts [4]. Methodologies were also framed wherein interested groups can formulate a group and each of the members within these groups can share and communicate to each other and these groups are called specific groups. The main advantage of these formation of groups is that the most relevant information needed by a group members can be retrieved from the members within the group or from the specific designated groups[5][6]. This popularity resulted in the formulation of social networking groups such as Twitter, Orkut , Face book [7][8].

Today, fan page is one amongst the different groups available in the social media, through this page important conversation and interesting communications are broadcasted among these groups. This fan page can be very much useful to advertise a particular brand, the brand advertisements in Face book dominate the Twitter and YouTube [1].This indirectly resulted in privacy and secrecy concerns about misusing the crucial information by internet users, one primary concern is that, virus authors use this social networking medium as a base and transmitting the virus among the groups [8]. Of late, Face book scams shoot the news with the tremendous increase in the number of fraud cases. This forced people rethink about the privacy of their Face book profiles, and also confirmed that several apps used in the Face book are being shared among the unauthorized people.

A. Privacy Issues in Social Media

There are several issues with regard to privacy, such as user Anonymity, where the identity of a user is exposed by the attacker and tries to at victim's profile. In De-Anonymization attack, the group member ship information is hacked by the attackers and tries to send anonymous mails by the attackers from the victims group. In the neighborhood attack, the neighbors around the network try to attack the victims. [5]

Apart from the privacy issues, there are several other issues, like user Profile leakage, leakage of information to third parties and Profile cloning.

Several other issues, that are target towards the privacy, include Spam mail attacks on Emails, Broad cast spam attacks, context spam attacks, where the attackers attack the victims sites, by sending several bulk mails.[5] . These are the factors that cause problems with regard to the privacy among the social networking sites.

It is therefore necessary to upgrade the present security methods of the face book and explore the privacy methodologies related to Social Networking Sites. In this paper, we have analyzed the offered privacy methodologies for the social networking sites, and in particular focused on the De-Anonymization attack and propose some new privacy model to strengthen the existing ones. To overcome these disadvantages privacy preserving approach together with network security approaches have been listed the literature [9][10]. In most of these approaches the authors have considered only about presenting the sensitive information and very little work is reported about the sharing of the information together with protection of the sensitive information.

In this article, we proceed to describe the methodology wherein each of the group members within a group are associated with a tag and for the efficient retrievals , the related tag of the images are given as input to the Generalized Gaussian Mixture Model for the experimentation purpose we have considered dataset of Flicker.

Each of the images in the dataset are subjected to normalization varying the invariable rotation are overcome. For this purpose the concept of Local Binary Pattern is used. Each of the user registered within the group are given an id and a code book is generated by summing up ids of all the individuals within the group. This sum is considered as the group code and with this as a tag, the images are retrieved. The main advantage is that if an unauthorized person wants to access the data from the group, the group id is to be understood. Therefore by this methodology we can overcome the fore said disadvantages. The rest of the paper is organized as follows. Section 2 of the paper deals with probability density function of the considered generalized GMM. In section 3 of the paper the details of the dataset are considered and presented. Section 4 of the paper deals with procedure of normalization based on a Local Binary Pattern. In section 5 of the paper the feature extraction is presented by using the concepts of relevance score and tagging. The results derived together with experimentation are presented in section 6. The conclusion is highlighted in section 7 of the paper. The future scope is presented in section 8 of the paper.

B. Related Work

Many authors have discussed about the issues of privacy in social networks. Most of the works are based on using Anonymization techniques (Zhou et.al(2008))[22],Encryption based(Guha et.al(2008))[23], Optimization model base, collaborative technique based(Blosser et.al(2008))[24], K-anonymity and sensitivity based approach(Ford et.al(2009))[25], Anonymized graphs(Narayanan et.al(2009))[26] access control model (Fong et.al(2009))[27]. Tang et.al(2010)[28] proposed a model using K-nearest neighborhood along with EBB algorithm for utilizing the privacy and the concepts of sub-graphs are considered by(Lan et.al(2010))[29] and the concepts of friendship means have been focused to ensure privacy and sending the sensitive information to these links was proposed by Heathely et.al(2013)[30].

In most of these models, the authors have proposed models towards the usage of models based on K-anonymity, L-diversity and the main disadvantages with these models is that they are prone to be a loss of information also Anonymization techniques failed to preserve the data based on dynamic releases. Further to add the models based on distributed approaches failed to uphold the privacy since preserving the privacy in a network environment in these cases are most sensible. To add most of the works proposed by the earlier researches are subjected to model the privacy issues which failed to overcome the attacks like homogeneity attack, background knowledge attack, distance based attacks and sensitive attacks. To overcome these advantages, the proposed article presents a model varying the disadvantages cited about can be overcome. Since it is a model based approach it holds the issues of homogeneity since every group which is subjected to the model presented in section 2 generates a unique PDF and also the model is more robust since it has a shape parameter and scale parameter which helps to model different sizes of groups and generating unique PDF's to each of these proposed groups. By varying the shape parameter and scale parameter, the model can be further extended for partitioning the groups into subsets where each subset can be a related or non-related.

II. Generalized Gaussian Mixture Model

In this article Generalized Gaussian Mixture Model is used to classify the images more appropriate basing on the symmetry of distribution. The probability density function of the Generalized Gaussian Mixture Model is presented in the following equation

$$f(z \mid \mu, \sigma, P) = \frac{1}{2\Gamma(1+\frac{1}{P})A(P,\sigma)} e^{-\left|\frac{(Z_i-\mu_i)}{A(P,\sigma)}\right|^P} \quad \text{---(1)}$$

$$\sigma > 0, A(P,\sigma) = \left[\frac{\sigma^2\Gamma(1/P)}{\Gamma(3/P)}\right]^{\frac{1}{2}} \quad --(2)$$

Where μ and σ are the mean and standard deviation and Γ defines the General Gamma Variate.

III. Dataset Considered

To perform the experiment we have considered the Flick database consisting of 25,000 images and each image is labeled with a tag description with labels like nature, cigar, flora and watch etc. Each image is coupled with a tag depiction; among these images 450 have unique tags. The experimentation is performed by taking into account 100 images, Query image is considered with the size 100 x 100.

IV. Local Binary Pattern

Local Binary Pattern is used to encode the relationship between the reference pixels with its surrounding neighbors by computing gray-level values. The Local Binary Pattern value is computed by comparing gray-scale value with its neighborhood.

$$\text{LBP}=\sum_{p=1}^{P})P-1)X\, f_1(l(g_p)-I(g_c))$$

$f_1(x) = 1 \quad x \geq 0$

$= 0 \quad$ otherwise --(3)

Fig. 1. Dataset considered from Flicker

V. Feature Extraction

For mining the images effectively from the database, Feature plays a crucial part. The features may perhaps be low level features or high level features which can be correlation, size, moments, histogram etc. Nevertheless, to retrieve the images more successfully, these features are to be linked with semantic understanding. The semantic interpretations aid to mine the data by means of the semantic characteristics and also lessen the semantic gap. These semantic traits are easily understood by the users when compared to the low level features which embrace contrast, symmetry, homogeneity and uniformity.

A. Score Level Fusion

For efficient image retrievals, score level fusion is used, the procedure operates on a Logical AND/ OR operation, where the relevancy is indicated as 'Y', and non-relevancy by 'N'.

B. KL-divergence

KL-divergence is used for the purpose to measure the distance between two probability density functions. It is a non-symmetric measure of the differences between two probability distributions .It is also known as relative entropy and information divergence.

$$KL(p_1, p_2) = \int p_1(x) \log\left(\frac{p_1(x)}{p_2(x)}\right) dx \quad --(4)$$

Where 'p1', 'p2' are the two Probability Density Functions

VI. Experimentation

In this model each of the users are identified basing on the grouping interest. These related users are formatted into a group by registering themselves with the data consisting of their e-mail ids and the group interest is considered and is tagged. These tags along with the e-mail ids are fused using score level fusion, discussed in above section-5.1 and these fused values are given as inputs to the model depicted in section 2 of the paper. In order to transmit or communicate the information among the groups, the authentication is to be established. And for the identification of the relatedness, the PDF's are compared using KL-divergence, proposed in section-5.2 and the authentication users are allowed to communicate and share the information.

VII. Conclusion

In this article a new framework is proposed to uphold the privacy issues by proposing the new methodology based on Generalized Gaussian Mixture model. The methodology developed helps in safeguarding the privacy of the information shared among the groups, such that the users can share the data with great deal of confidence. This method will be useful in Social Networking Sites and in particular on the Face book.

VIII. Future Scope

In this paper, a methodology is presented to safe guard the group user's information in the social networking media. This paper address the methodology for overcoming the De-Annomyzation attack, however , effective mechanism are to be developed to overcome the other attacks , such as spam attack, hijacker attack . Also, most of the data can be transmitted by using the watermarking techniques. Therefore methods are to be developed to overcome this issue.

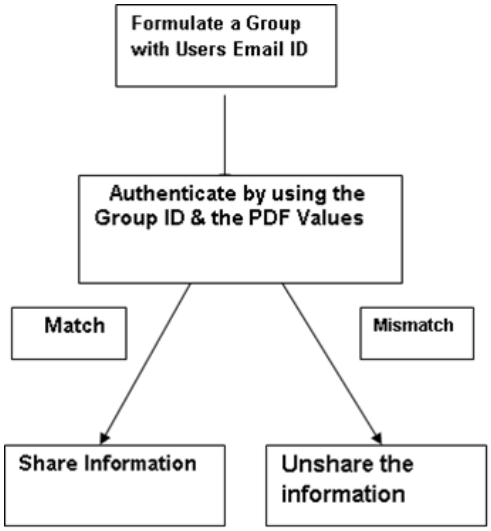

Fig. 2. The proposed architecture

REFERENCES

[1] MARK Hachman (Aoril 23,2012). "Facebook Now Totals 1.20 billion Users, Profits Slip". PCMag.com. Retrieved September 24, 2013.

[2] D. M. Boyd and N. B. Ellison, "Social Network Sites: Definition, History, and Scholarship,"J. Comp.-Mediated Communication., vol. 13, no. 1, Oct. 2007, pp.210–30.

[3] S. B. Barnes, "A Privacy Paradox: Social Networking in the United States,"First Monday , vol. 11, no. 9, Sept. 2006.

[4] Aimeur, E.; gambus,S.; Ai Ho; , "UPP: User Privacy Policy for Social Networking Sites,"Internet and Web Applications and Services,2009. ICIW '09.Fourth International Conference on vol., no., pp.267-272,24-28 May 2009.

[5] A Survey of Privacy and Security Issues Social Networks http://www.cse.wustl.edu/~jain/cse571-11/ftp/social/index.htm

[6] Stutzman, F.and Kramer-Duffield, J. Friends only: Examining a Privacy Enhancing behaviour in Facebook. In Proc. CHI'10.ACM Press, 2010.1553—1562.

[7] A. Ho, A4. Maiga, and E. Aimeur, "Privacy protection issues in social networking sites,"IEEE/Acs International Conference on Computer Systems and Applications 2009 (AICCSA 2009),PP.271-278,Country,2009

[8] Thomas,K., Grier, C., and Nicol,D.M.unFriendly: Multi party privacy risks in social networks.In proceedings of the 10th international conference on Privacy enhancing technologies(2010),Soringer-Verlag,pp.236

[9] http://www.facebook.com\privacy

[10] Ai Ho; Maiga, A.; Aimeur, E.; , "Privacy protection issues in social networking sites," Computer Systems and Applications, 2009. AICCSA 2009.IEEE/ACS International Conference on vol.,no., pp.271-278, 10-13 May 2009

[11] Xi Chen; Shuo Shi; ,"A Literature review of Privacy Research on Social Network Sites," Multimedia Information Networking and Security,2009.MINES'09.International Conference on, vol.1,no.,pp.93-97,18-20 Nov.2009

[12] SeyedHossein Mohtasebi and Ali Dehghantanha," A Mitigation Approach to the Malwares Threats of Social Network Services," Muktimedia Information Networking and Security,2009. MINES'09. International Conference on, vol.1,no.,pp.448-459,2011

[13] Mohammad Mannan, Paul C. Van Oorschot," privacy-Enhanced Sharing of Personal Content on the Web," Security And Privacy- Misc" , pp.487-496,April 21-25,2008 Beijing, China

[14] Privacy Policy Facebook (2011), www.facebook.com/policy.php

[15] Chi Zhang; Jinyuan Sun;, "Privacy and Security for Online Social networks:Challenges and Opportunities," IEEE Network,Aug.2010

[16] Vorakulpipat, Marks, Rezgui, " Security and Privacy Issues in Social Networking Sites from User's Viewpoint,"IEEE Network,Jun.2011

[17] P.Kodeswaran, and E.Viegas, "Towards A privacy preserving Policy Based Infrastructure for Social Data Access To enable Scientific Research,"2010 Eighth Annual International Conference on Privacy,Security and Trust,Jun 2010

[18] I. Polakis and G. Kontaxis," Using Social Networks to harvest Email Addresses," In Proc.CHI"10. ACM Press,2010

[19] C.Squicciarini and M.Shehab," Privacy policies for shared content in social network sites," In Proc.Chi'10.Acm Press,30 June 2010

[20] Yabing Liu and P. Gummadi, "Analyzing Facebook Privacy Settings:User Expectations vs. Reality," In Proceedings of the 10th international conference on Privacy enhancing technologies(2011)

[21] P. Joshi and C Kuo ," Security and Privacy in Online Social Networks: A Survey",IEEE Network,2011

[22] B. Zhou, Jian Pei,Wo-Shun Luk, " A brief survey on anonymization techniques for privacy preserving publishing of social network data," ACM SIGKDD Explorations Newsletter, Vol. 10,pp. 12-22,2008.

[23] Saikat Guha , Kevin Tang, Paul Francis, "NOYB: Privacy in Online Social Networks", In Proc. of first workshop on Online social networks WOSN'08, ACM New YORK, NY, USA, pp 49-54,2008.

[24] Gary Blosser, Justin Zhan, "Privacy Preserving Collaborative Social Network", In Proc. Of International Conference on Information Security and Assurance ISA 2008,Busan,pp.543-548,2008.

[25] Roy Ford, Traian Marius Truta, and Alina Campan, "P-Sensitive K-Anonymity for Social Networks".

[26] A. Narayanan, V. Shmatikov, "De-anonymyzing social networks", In Proc of 30th IEEE Symposium on Security and Privacy, Berkely, CA, pp 173-187,2009.

[27] Philip W. L. Fong, Mohd Anwar, and Zhen Zhao," A Privacy Preservation Model for Facebook-style Social Network Systems", In: Computer Security- ESORICS 2009, Lecture Notes in Computer Science, Vol.5789,2009,pp 303-320,2009.

[28] X. Tang and C. C. Yang, " Generalizing Terrorist Social Networks with K-Nearest Neighbor and Edge Betweenness for Social Network Integration and Privacy Preservation," In Proc. of IEEE International Conference on Intelligence and Security Informatics, 2010.

[29] Lihui Lan, Shiguang Ju Hua Jin,"Anonymizing Social Network using Bipartite Graph",In Proc. of International Conference on Computational and Informatics Sciences(ICCIS), Chengdu, pp 993-996,2010.

[30] Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham, "Preventing Private Information Inference Attacks on Social Networks", In: IEEE Transactions On Knowledge And Data Engineering, Vol.25, No.8,pp 1849-1862,2013.