IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 01-08

Preventing Profile Sensitive Information Inference Attacks in OSN using Data Sanitization Technique

Ms. Priti N. Rathod[1], Prof. M. B. Kalkumbe[2]
[1]*ME Student: Computer Science & Engineering Dept. MSSCET, Jalana, Marathwada University, Aurangabad, India.*
[2]*Professor: Computer Science & Engineering Dept. MSSCET, Jalana, Marathwada University, Aurangabad, India.*

Abstract: *(OSN) On-line social networks like Facebook are increasingly utilized by many people. These networks enable users to publish their own details and modify them to contact their friends. The information generated by social media services typically said because the social network data. In several things, the information has to be revealed and shared with others. Social networks ar on-line applications that enable their users to attach by means that of assorted link varieties. Some of the knowledge disclosed within these networks is supposed to be personal. A privacy breach occurs when sensitive information about the user, the information that an individual wants to keep from public, is disclosed to an adversary. Private info discharge may be a crucial issue in some cases that is termed inference Attack. In this research paper, the proposed system tries to hide Personal Information (PI) of user automatically, at the time of account creation. To protect against inference attacks, analysis propose an information sanitation methodology jointly manipulating user profile and relationship relations. In this methodology, the most challenge of protection of Sensitive personal info is self-addressed. In this research we propose a Method that takes advantages of various data manipulating methods and guarantee maximum protection to personal information of Social media users. Flexibility of System in terms of filtering options is enhanced Proposed system can work reasonably to balance privacy and data utility. Third party users cannot obtain necessary information to accurately predict sensitive information. Reduce the chance of attacks due to the secure communication.*

Keywords: *Client/server, inference attack, Distributed Systems, hacking, phreaking*.

I. Introduction (Heading 1)

Even though online social networking sites ensure the protection of data, there are many incidents regarding the private information drip of these social networking sites. Many on-line social network (OSN) owners frequently publish information collected from their users' on-line activities to 3rd parties like sociologists or business corporations. These third parties any mine the information and extract data to serve their various functions. In the method of business information to those third parties, network owners face a nontrivial challenge: how to preserve users' privacy while keeping the information useful to third parties. Failure to protect users' privacy may result in severely undermining the popularity of OSNs as well as restricting the amount of data that the OSN owners are willing to share with third parties. Although these OSN provide various features to interact with the people they lack security features. The main advantage of these online social networks other than communication is the marketing and research fields. OSN now become a medium of major business center. Many companies can post ads on these Websites and gain a huge profit. By using data mining algorithms on these networks able to analyze data and arriving on conclusions. These predictions and conclusion may violate others privacy. In OSNs, info filtering can even be used for a distinct, more sensitive, purpose. This is because of the very fact that in OSNs there's the chance of posting or commenting alternative posts on specific public/private areas, called in general walls. Information filtering will so be accustomed provide users the power to mechanically management the messages written on their own walls, by filtering out unwanted messages. It believes that this is a key OSN service that has not been provided so far. Indeed, nowadays OSNs give little or no support to stop unwanted messages on user walls.

II. Literature Survey

A. Anonymization and De-anonymization

Privacy is typically protected by anonymization methods, i.e., removing information regarding name, religion, political view, etc. However, such network could be de-anonymized by utilizing background knowledge such as reference network. For example, De-anonymized approaches utilize 'network mapping' to map nodes from reference network to anonymized network.

In [3], the authors propose a community-enhanced de-anonymization approach to re-identify users, a divide-and-conquer approach to strengthen the power of such algorithms. Our approach partitions the networks into communities' and performs a two-stage mapping which first partitions the network into communities and then carries out a two-stage mapping: first mapping communities then the entire network.

In [4], the authors consider a de-anonymization algorithm to re-identify the users in an anonymized social network based on network topology, namely, mapping the anonymous target graph and the aggregated graph from multiple social networks.

B. Inference Attacks and Protecting Methods.

Similarly, the work in [6] indicates that users' sensitive information can be inferred based on friendship information and group memberships, and it also shows that disclosure of one user's hidden attribute would breach her friends privacy.

III. Goal And Objectives

A malicious/third party user will try to infer the private information which will lead to security issues[8]. Unknown user taking our information without our knowledge is called inference attack. Inference attack is not only privacy violation, it will also lead to other issues like identity theft and phishing attack. The main objective is to develop the system is growing rate of crimes on Internet, for protection of sensitive data from hackers and other third party applications, for preserving the users privacy and providing the secure communication on online social network

In propose a method of measuring the amount of information that a user publish on online social network and which automatically determines which information (on a per-user basis) should be removed to increase the privacy of an individuals. The planned system provided additional protection to non-public information by hiding it throughout account creation. This has will increase the general security of social media information. And the accessibility of data to 3rd party users is narrowed down.

IV. Planned System

In this paper, we've got a bent to focus on latent-data privacy. we've got an inclination to assume third party users may collect anonymous user info from social networks. Some users disclose their sensitive information, whereas others do not. However, third party users can do de-anonymization actions and a lot of infer sensitive information of users. we have a tendency to initial investigate the simplest way to infer sensitive information hidden inside the free info. Then, we have a tendency to propose some effective info sanitisation strategies to stop information AN inference attacks. On the alternative hand, the sanitized info obtained by these strategies should not reduce the pricey profit brought by the pr info resources, so as that non-sensitive information can still be inferred and utilized by third party users. To launch associate degree inference attack by third party users, we have a tendency to use a typical inference attack, spoken as collective inference, as a case study. And sensitive information hidden by using NPL and in addition audio will play for request to user i.e. text to speech .We gift a totally distinctive implementation technique for collective logical thinking. Collective inference within the main suppose iteratively propagating current predicting results throughout a network to boost prediction accuracy, so we'd wish to think about the simplest way to best predict sensitive information in each iteration. The planned methodology provided a secure communication to the users in on-line social network (OSN). By filtering the undesirable content by using language process (NLP), NLP provides the quantity that increments the worth of Vulgar perennial words within the information base table.

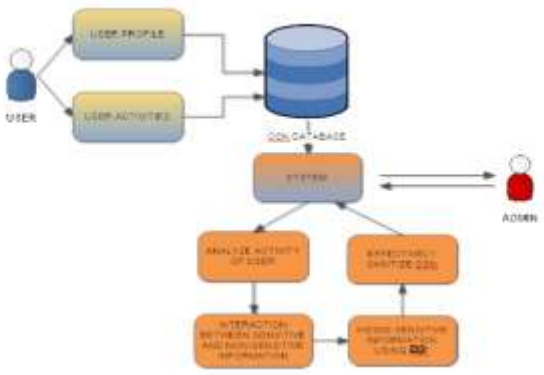


Fig. 1: Planned System style

C. Working of Planned System

In this [Fig. 1] the user can register to the system with traditional info. At the time of registration the OSN system can hide the user's sensitive info. For login to the system, user can enter the Username and password, if entered details ar correct then the system can direct him to home page otherwise it'll shows miscalculation message. once Login, User can share the post, Post the status, set the setting to profiles. Send the messages to different users by checking the attributes.

User can perform the User Attribute like profile setting, post sharing, like or comment onto the post, posting text message, text documents, video and message causing to the another users by matching the attributes. In OSN System, The OSN system: Check sensitive and non-sensitive info of all users Check the all registered users sensitive info. It keep the sensitive attributes. The OSN can give the privacy for users like and comments posts. concealment sensitive info using NLP. Text to speech converter additionally for notification used.

When a brand new account is formed on Facebook, the private sensitive info is offered to all or any different users. once the account is formed using planned system, this info is automatically hidden from different users. This info would be created out there only once friend request acceptance. OSNs permit their users to manage and manage privacy settings on their profile usually configuring these settings for each item is confusing and a time intense task. the opposite is regarding communication strategy.

In the existing system, once user creates a brand new account on Facebook, personal info of his/her account (Work, Education, different Basic information) is visible to all or any the users present on Facebook. Anyone, who isn't within the Friend's list of a user, are ready to see these professional and private details. A security issue happens once a hacker gains unauthorized access to private sensitive info of user. Privacy problems, those involving the unwarranted access of personal info, don't essentially got to involve security breaches. somebody will gain access to private information by only looking user on social network sites.
To overcome this security issue, the planned system provides tight security as below:
• Once a brand new user creates AN account on Facebook, all his professional and private details ar Hidden by the system.
• If somebody already on Facebook searches for this freshly created account, existing user wouldn't be ready to see new user's details.
• Once new user is added to the Friend list of existing user, all his details would then be visible to existing user. it'll not be accessible to the 'Public' class of user.

D. Filtering Unwanted Messages

This is the primary proposal of a system to automatically filter unwanted messages from OSN user walls on the idea of message content. The Naive Bayes Classifier is predicated on the bag-of-words model. With the bag-of-words model we have a tendency to check that word of the text-document seems in a very positive-words-list or a negative-words-list. If the word seems in a very positive-words-list the overall score of the text is updated with +1 and the other way around. If at the end the overall score is positive, the text is classed as positive and if it's negative, the text is classed as negative.
NLP (NATURAL LANGUAGE PROCESSING)

Natural-language method (NLP) could be a neighbourhood of subject area and computing involved the interactions between computers and human (natural) languages, particularly the simplest way to program computers to fruitfully methodology big amounts of language information.

E. NLP algorithmic rule for text filtering

Input: Sentence

Step 1: Collect input text information
String s="am is hot ar java was";

Step 2: Preprocessing Document
Preprocessing document consists of following steps-
Tokenization, stemming, case transformation, stop word removal filtering etc

Step 3: Implementing Naive Bayes Text Classification
To implement a Naive Bayes Text Classifier we've got a text category (Naive Bayes also can be used to classify non-text / numerical datasets.)

We have a Naive Bayes Text category, that accepts the input values for stopWrds and stopWrds1 as parameters for the "train()" methodology.
Then place the complete words one by one in new file that's in E://content.txt file. Then compare every word of E://content.txt with E:\\stopwords1.txt file.
As we are able to see, the training of the Naive Bayes Classifier is completed by iterating through all of the documents within the training set from all of the documents

Step 4: Check in Bags of words
- Then compare each word of E:\\content.txt with E:\\stopwords1.txt file.
Step 5: Filtering unwanted words using Sentiment Analysis with the Naive Bayes Classifier
- After pre-processing put remaining words one by one in content1.txt file using java language scanner function.
- Content1.txt file check these words in stopword1.txt file dataset whether it is present in dataset or not.
Output
Check the sentiment of message if it is positive posted message on user wall. If it is negative then message not posted on user wall.

V. Advantages Of Planned System

1. Flexibility of System in terms of filtering options is enhanced.
2. When a new user creates an account on Facebook, all his sensitive personal details are Hidden by the system
3. When someone already on Facebook searches for this newly created account, existing user would not be able to see new user's details
4. Proposed system can work reasonably to balance privacy and data utility.
5. Third party users cannot obtain necessary information to accurately predict sensitive information.
6. Reduce the chance of attacks due to the secure communication.

VI. Data Sanitization

We propose some effective data sanitization strategies to prevent personal information inference attacks. On the opposite hand, the sanitized information obtained by these methods shouldn't scale back the dear profit brought by the pr information resources, in order that non-sensitive info will still be inferred and utilized by third party users. To launch an illation attack by third party users, we use a typical illation attack, referred to as collective illation, as a case study. We gift a completely unique implementation technique for collective illation. Collective illation principally place confidence in iteratively propagating current predicting results throughout a network to boost prediction accuracy, thus we need to consider how to best predict sensitive information in each repetition [6]. As said before, when another record is made on Facebook, the individual touchy data is accessible to every other client. At the point when the record is made utilizing proposed framework, this data is naturally avoided other clients. Underneath screen shot [Fig.3] shows that the individual data of client, present in 'About' tab, won't be unmistakable to other existing clients. This data would be made accessible simply after companion asks acknowledgment. The hidden data are shown as dotted, suggests that the user seeing this data isn't allowed to scan it. These dotted lines would get replaced by actual data text only the Friend request is accepted by new user.

❖ **Profile Data sanitization Steps**:
Step1: Rad profile data from Database.
Step2: Sort profile database by sensitivity value.
Step3: Using collective Data sanitization algorithm Sanitize/Masking Sensitive attributes of profile.
Step4: Hiding the sensitive information from user profile during account creation.

FIG.2: Profile Hidden Information seen by user

VII. Result Analysis

A. Login page

After finish your registration then you can login in the web Page. A login is the entering of identifier information into as system by a user in order to access that system.

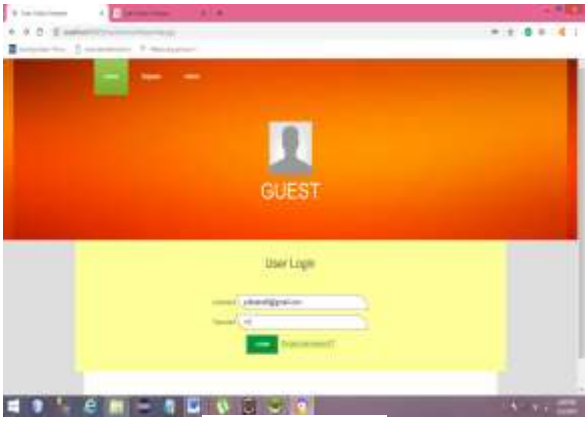


Fig 3 Login page

B. Home page

After Entering the login page you can check your activities, friends and you can manage your profile details by own setting and it would be automatically saved in your database.

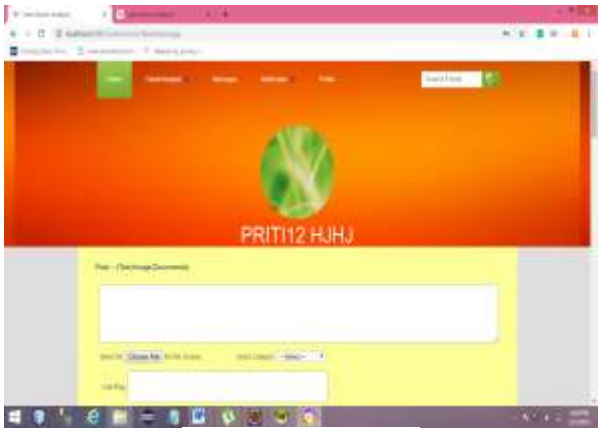


Fig 4 User home page

C. Admin page

The Administration Panel can be accessed, Using the first thing you will see is a prompt to login to access the administration panel. This username and Password is different from your sites login details. After you Login, you will be taken to the dashboard which will show you statistics of your site.

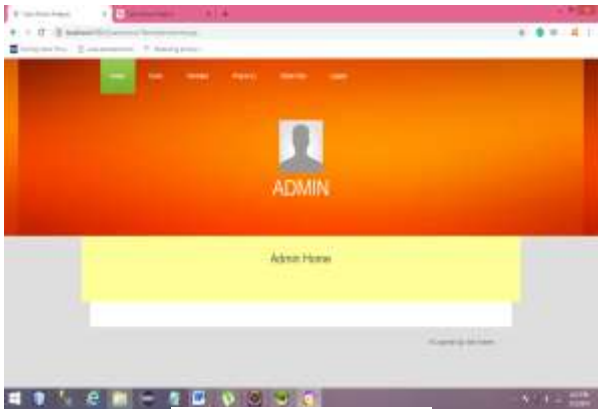


Fig 5 Admin home page

If you were login your admin you can see lot of stuffs, like Raw Profile, Raw Activities and so on. First thing you can see the Raw dataset to all the basic information of the User.

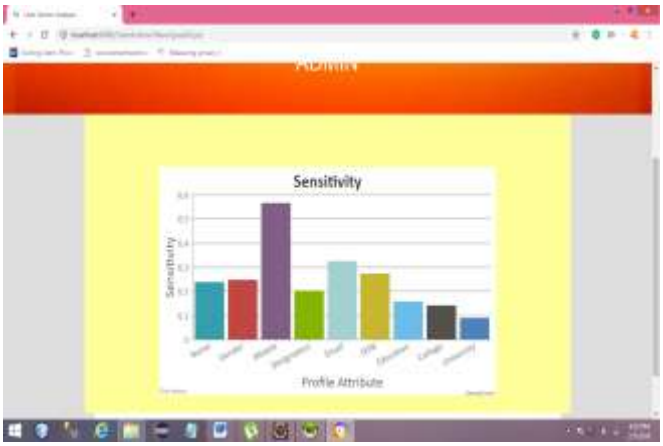
Fig 6. Raw dataset to all the basic information of the User

Fig 7 Raw Activities of all the Users

D. Sensitivity of profile attributes

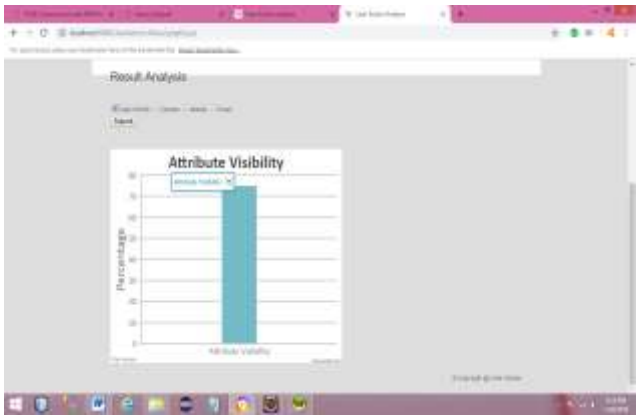
On social networking sites, there are far more attributes that affect user privacy, Sensitivity shows the risk associated with the attributes of the user. When the sensitivity of an attribute increases, the risk posed by information disclosure of the individuals also increases. Linkage Xuefeng Li, Yixian Yang, [9] calculated the sensitivity score for 11 attributes. His results indicated that the most sensitive attributes are related to phone

number, address, email, username, birthdate. In contrast, job details, relationship status, interest and education are not that sensitive for the users. For attribute visibility, we are using sensitivity values derived by Linkage Xuefeng Li , Yixian Yang.[9]. In real life, due to the complexity of settings, the actual settings are not consistent with a user's expectation, which means that using the profile setting cannot reflect the actual sensitivity of attributes.



Graph 1 Profile Attribute Sensitivity Graph

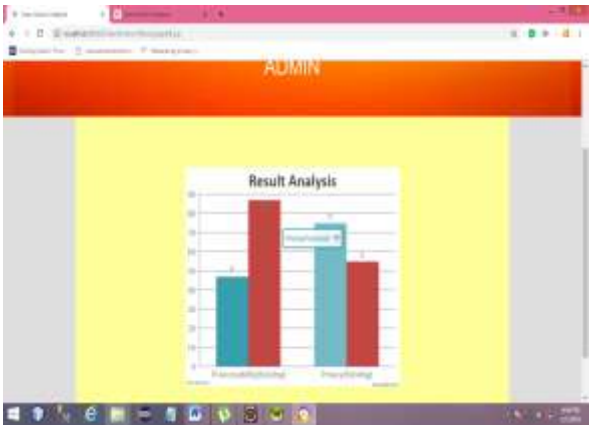
Visibility determines how widely accessible the attributes of a user are in an online social network This is means that if we are sharing more and more sensitive attributes privacy on user profiles get decreased. Hence, sharing information to such nodes could result in an unwanted privacy breach and became a chances of the sensitive information being leaked through them is high.



Graph2. Privacy on visibility of sensitive attributes

Following figure shows comparison of Existing and proposed system by assuming 50 people data. The comparison is done on two parameters of security; Accessibility of Personal information on Social sites and Security of Personal Information on Social sites.

As per research done for Social sites, the existing system provide 87% accessibility to Personal information of user. This can be misused by hackers. Proposed system tried sanitize the private info at the time of account creation. This has reduced the Accessibility to third party from87% to 47%. Also, the privacy of existing Social systems will be said to be concerning 55% as per social media survey report. The planned system has increased the safety level up to 75% by sanitizing sensitive info and vulgar words.



Graph.3. Comparison chart for existing security system and proposed system for social

VIII. Conclusion

As per analysis performed for this paper, we observed that the existing social networking sites allow display of Sensitive Personal information to users. A security concern may occur when a hacker gains unauthorized access to this information. Privacy problems, those involving the unwarranted access of private information, don't necessarily have to involve security breaches. Someone will gain access to private information by just adding an individual to friend list. To overcome this disadvantage of existing system, the proposed system provided more protection to personal data by hiding it during account creation. This has will increase the overall security of social media information. And the accessibility of personal information to 3rd party users is narrowed down. Lastly we conclude that with this work we hope to motivate advanced research in the field of data privacy specifically in the area of measuring user privacy, enabling selective sharing of sensitive data and protecting sensitive data from inference attacks in OSNs

In future, we would like to explore further generalization of the privacy scoring framework considering user's perspectives about the sensitivity of their data, to use Natural Language Processing paradigms and decide on that of the private info will be made accessible and that a part of the PI should be hidden at the time of account creation. As an extension to our work the proposed recommender system could be made more robust where it can filter out the sensitive contents like users' personal photos and videos from the general content before computing the privacy quotient.

Acknowledgment

I would like to take this opportunity to thank MSS trust and Principal DR. S. K. Biradar , MSSCET for giving me an opportunity to carry out my project work. I express my deep sense of gratitude towards Prof. G. P. Chakote, Head of the Department of Computer for his valuable guidance and encouragement. I wish to acknowledge my extreme gratitude to my guide Prof. M. B. Kalkumbe Mam for guiding me throughout the work on project. I have been greatly benefited by his valuable suggestion and ideas.

References

[1]. j. he, w. chu, and v. liu(2006), "Inferring Privacy Information from Social Networks," Proc. Intelligence and Security Informatics.
[2]. E. Zheleva And L. Getoor(2008), "Preserving The Privacy Of Sensitive Relationships In Graph Data," Proc. First AcmSigkdd Int'l Conf. Privacy, Security, And Trust In Kdd, Pp. 153-171.
[3]. S. Nilizadeh, A. Kapadia, and Y.-Y.Ahn, "Community-enhanced de-anonymization of online social networks," in Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, ser. CCS '14. New York, NY, USA: ACM, 2014, pp. 537– 548.
[4]. A. Narayanan and V. Shmatikov, "De-anonymizing social networks," in Proceedings of the 2009 30th IEEE Symposium on Security and Privacy, ser. SP '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 173–187.
[5]. B. Zhou, J. Pei, and W. Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data," SIGKDD Explor.Newsl., vol. 10, no. 2, pp. 12–22, Dec. 2008.
[6]. A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: Inferring user profiles in online social networks," in Proceedings of the Third ACM International Conference on Web Search and Data Mining, ser. WSDM '10. New York, NY, USA: ACM, 2010, pp. 251–260.
[7]. J. K. Jonghyuk Song, Jonghyuk Song, ―Inference attack on browsing history of twitter users using public click analytics and twitter metadata,‖ IEEE Transactions on Dependable and Secure Computing, 2014.
[8]. Ahmadinejad SH, Fong PW (2013)" On the feasibility of inference attacks by third-party extensions to social network systems", Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security, ASIACCS, pp. 161-166.
[9]. A Privacy Measurement Framework for Multiple Online Social Networks against Social Identity Linkage Xuefeng Li 1,2,* , Yixian Yang 1,2, Yuling Chen 2 and Xinxin Niu 1,2.