ISSN: 0976-3104
SPECIAL ISSUE: COMPUTER SCIENCE

ARTICLE

COLLECTIVE DATA-SANITIZATION FOR PERSONAL INFORMATION PROTECTION

Pranjali Kothawade*, Suhas Patil

Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University), College of Engineering, Pune, INDIA

ABSTRACT

Background: *On-line social networks like Facebook are increasingly utilized by many people. These networks allow users to publish their own details and enable them to contact their friends. Some of the information revealed inside these networks is private. These networks allow users to publish details about themselves and to connect to their friends. Some of the information revealed inside these networks is meant to be private. 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.* ***Methods:*** *Private information leakage could be an important issue in some cases which is called Inference Attack. In this research paper, the proposed system tries to hide Personal Information of user automatically, at the time of account creation. To protect against inference attacks, research propose a data sanitization method collectively manipulating user profile and friendship relations.* ***Results:*** *In this methodology, the main challenge of protection of Sensitive personal information is addressed.* ***Conclusions****: 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.*

KEY WORDS
Online Social Networks (OS Ns), Collective Inference, Data Sanitization, Inference attacks.

Received: 14 May 2018
Accepted: 7 June 2018
Published: 10 June 2019

***Corresponding Author**
Email:
pranjali85bahalkar@gmail.com
Tel.: +91-9503701844

INTRODUCTION

Social media

A social networking service (also social networking site, SNS or social media) is an online platform that people use to build **social networks** or **social relations** with other people who share similar personal or career interests, activities, backgrounds or real-life connections. The variety of stand-alone and built-in social networking services currently available online introduces challenges of definition; however, some common features exist microorganisms [1].

Voids of social media

The rapid growth and ubiquity of online social media services has given an impact to the way people interact with each other. Online social networking has become one of the most popular activities on the web. Social network analysis has been a key technique in modern sociology, geography, economics, and information science. The data generated by social media services often referred to as the social network data. In many situations, the data needs to be published and shared with others.

Social networks are online applications that allow their users to connect by means of various link types. As part of their professional network; because of users specify details which are related to their professional life. These sites gather extensive personal information, social network application providers have a rare opportunity direct use of this information could be useful to advertisers for direct marketing. Publish data for others to analyze, even though it may create severe privacy threats, or they can withhold data because of privacy concerns, even though that makes the analysis impossible [2].

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. For examples, business companies are analyzing the social connections in social network data to uncover customer relationship that can benefit their services and product sales. The analysis result of social network data is believed to potentially provide an alternative view of real-world phenomena due to the strong connection between the actors behind the network data and real world entities. Social-network data makes commerce much more profitable. On the

Other hand, the request to use the data can also come from third party applications embedded in the social media application itself.

For instance, Facebook has thousands of third party applications and the number is growing exponentially [3]. Even though the process of data sharing in this case is implicit, the data is indeed passed over from the data owner (service provider) to different party (the application) The data given to these applications is usual not sanitized to protect users' privacy. Desired use of data and individual privacy presents an opportunity for privacy-preserving social network data mining. That is, the discovery of information and relationships from social network data without violating privacy.

Privacy concerns in social networks can be mainly categorized into two types:

- Inherent-data privacy
- Latent data privacy

IIOAB JOURNAL
ISSN: 0976-3104

Inherent-data privacy is related to sensitive data contained in the data profile submitted by users in order to receive data-related services [4].

Communication strategy on social media

While a great amount of literature has focused on the relationship between communication strategies and corporate reputation, there is no systematic research on the different kinds of social media communication strategies. Based on the corporate reputation and social media literature, this paper aims to contribute to this gap in the research in two main ways.

- First identifying which social media communication strategy is more effective with contrasting levels of reputations [5];
- Second, analyzing the differences between high- and low-reputation companies with respect to their ability to use corporate communication.

MATERIALS AND METHODS

Data sanitization

We propose some effective data sanitization strategies to prevent information inference attacks. On the other hand, the sanitized data obtained by these strategies should not reduce the valuable benefit brought by the abundant data resources, so that non-sensitive information can still be inferred and utilized by third party users. To launch an inference attack by third party users, we employ a typical inference attack, called collective inference, as a case study. We present a novel implementation method for collective inference. Collective inference mainly rely on iteratively propagating current predicting results throughout a network to improve prediction accuracy, thus we need to consider how to best predict sensitive information in each repetition [6].

Working of this module:

Algorithm:

- User creates an Account on social media sites.
- It stored the sensitive attributes.
- All personal data is automatically hides in the database record.
- Any Third party users search this account that time they are not see user's personal sensitive data.
- When user accept a friend request for another user only that authorized users are see all personal sensitive information.
- When user provide the accessibility for personal data to friend list friends those users are only see and access the active users information.
- The OSN will provide the privacy for users like and comments posts.
  Data sanitization method provides the Accessibility and Security Feature.

NLP (Natural Language Processing)

Natural-language processing (NLP) is an area of computer science and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to fruitfully process large amounts of natural language data. Challenges in natural-language

Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output.

Processing frequently involve speech recognition, natural-language understanding, and natural-language generation. Many different classes of machine learning algorithms have been applied to natural-language processing tasks. These algorithms take as input a large set of "features" that are generated from the input data. Some of the earliest-used algorithms, such as decision trees, produced systems of hard if-then rules similar to the systems of hand-written rules that were then common. Increasingly, however, research has focused on statistical models, which make soft, probabilistic decisions based on attaching real-valued weights to each input feature. Such models have the advantage that they can express the relative certainty of many different possible answers rather than only one, producing more reliable results when such a model is included as a component of a larger system.

Algorithm:

- Hiding some sensitive information and
- Effective sanitize online social media network.
- NLP provides the Number that increments the value of Vulgar repeated words in the Data base table.

ISSN: 0976-3104

- To use Natural Language Processing paradigms and decide on which of the personal information can be made available and which part of the PI should be hidden at the time of account creation.
- Standard API's of NLP Used for implementation.

Text to speech convertor

A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech Synthesized speech can be created by Concatenating pieces of recorded speech that are stored in a database. Systems differ in the size of the stored speech units; a system that stores phones or diaphones provides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output.

Algorithm:

- Audio request for user when notifications as well as any comment, Friend request occur on social sites.
- When user login the Facebook and notifications and requests indications are click then user listen that time audio message.

Proposed Architecture

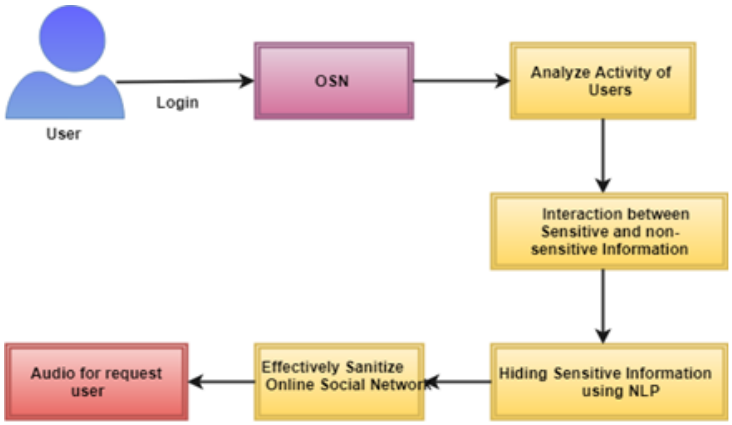

Fig. 1: Architectural diagram for proposed system.

In this [Fig. 1.] The user will register to the system with normal information.

At the time of registration the OSN system will hide the user's sensitive information. For login to the system, user will enter the Username and password, if entered details are correct then the system will redirect him to home page otherwise it will shows an error message.

After Login, User will share the post, Post the status, set the setting to profiles. Send the messages to other users by checking the attributes. User will perform the User Attribute like profile setting, post sharing, like or comment onto the post and message sending to the another users by matching the attributes.

In OSN System, The OSN system: Check sensitive and non-sensitive information of all users Check the all registered users sensitive information. It stored the sensitive attributes. The OSN will provide the privacy for users like and comments posts. Hiding information using NLP.Text to speech convertor also for notification used .

RESULTS

Final result of the implemented work shows that more security is now provided for the protection of personal information.

[Fig. 2] shows comparison of Existing and proposed system. The comparision is done on two parameters of security; Accessibility of Personal information on Social sites and Security of Personal Information on Social sites.

ISSN: 0976-3104

As per research done for Social sites, the existing system provide 84% accessibility to Personal information of user. This can be misused by hackers. Proposed system tried to hide the Personal Information at the time of account creation. This has reduced the Accessibility to third party from84% to 40%.

Also, the Security of existing Social systems can be said to be about 55% as per social media survey report. The proposed system has increased the security level up to 75% by hiding sensitive information and vulgar words.

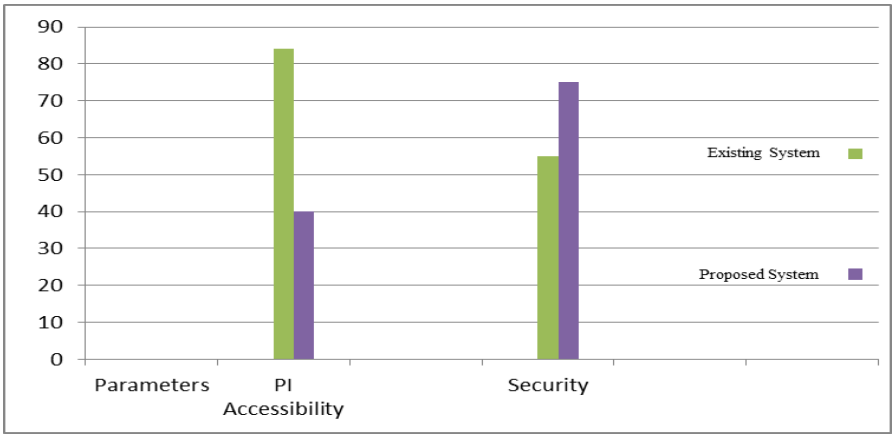

Fig. 2: Comparison chart for existing security system and proposed system for social media.

As said earlier, when a new account is created on Facebook, the personal sensitive information is available to all other users.

When the account is created using proposed system, this information is automatically hidden from other users.
Below screen shot [Fig. 3] displays that the personal information of user, present in 'About' tab, will not be visible to other existing users. This information would be made available only after friend request acceptance.

The hidden information will be shown as dotted, means the user seeing this information is not allowed to read it. These dotted lines would be replaced by actual information text only when the Friend request is accepted by new user.

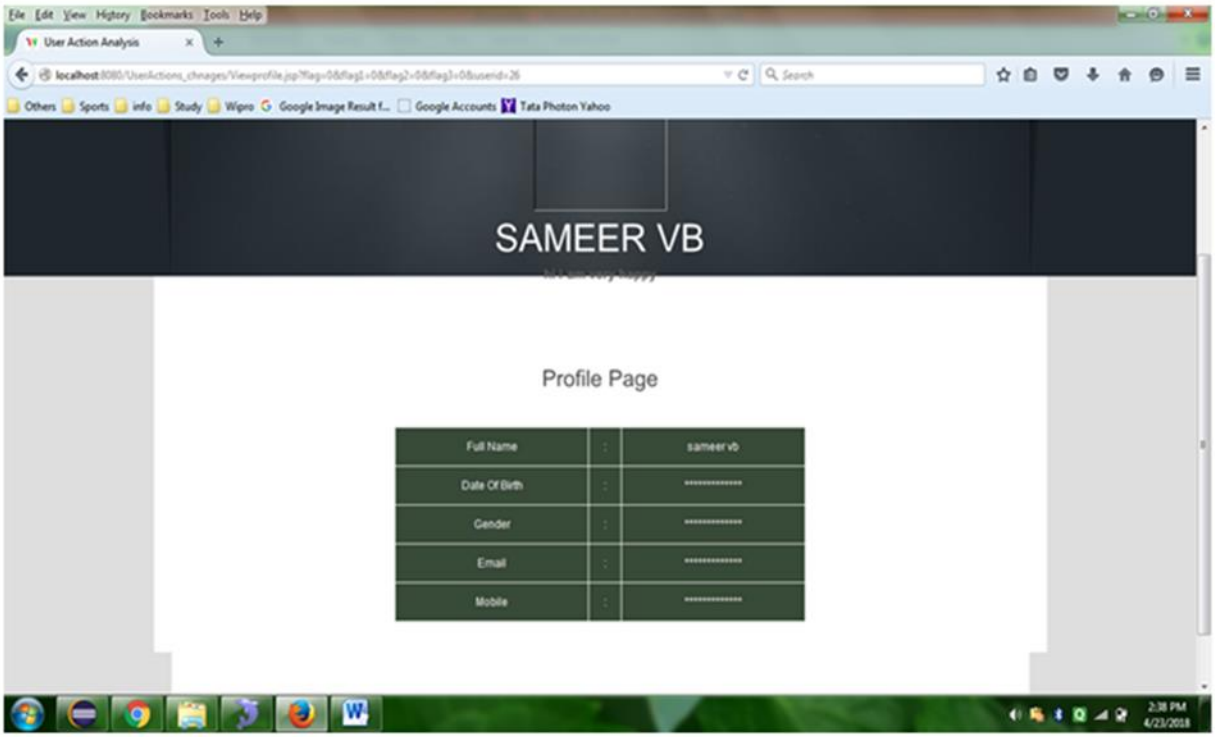

Fig. 3: Hidden Information seen by user

IIOAB JOURNAL
ISSN: 0976-3104

DISCUSSION

In the existing system, when user creates a new account on Facebook, personal information of his/her account (Work, Education, other Basic information) is visible to all the users present on Facebook. Anyone, who is not in the Friend's list of a user, will be able to see these professional and personal details. A security issue occurs when a hacker gains unauthorized access to personal sensitive information of user. Privacy issues, those involving the unwarranted access of private information, don't necessarily have to involve security breaches. Someone can gain access to confidential information by simply searching user on social network sites.

To overcome this security lacuna, the proposed system provides tight security as below;

- When a new user creates an account on Facebook, all his professional and personal details are Hidden by the system.
- When someone already on Facebook searches for this newly created account, existing user would not be able to see new user's details.
- When new user is added to the Friend list of existing user, all his details would then be visible to existing user. It will not be available to the 'Public' category of user.

CONCLUSION

As per research done 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 issues, those involving the unwarranted access of private information, don't necessarily have to involve security breaches. Someone can gain access to confidential information by simply adding a person to friend list.

To overcome this shortcoming of existing system, the proposed system provided more protection to personal data by hiding it during account creation. This has increases the overall security of social media data. And the accessibility of information to third party users is narrowed down.

The future scope for system can be; to use Natural Language Processing paradigms and decide on which of the personal information can be made available and which part of the PI should be hidden at the time of account creation.

CONFLICT OF INTEREST
None

ACKNOWLEDGEMENTS
None

FINANCIAL DISCLOSURE
None

REFERENCES

[1] J He, W Chu, V liu. [2006] Inferring Privacy Informatio from Social Networks, Proc. Intelligence and Security Informatics.
[2] E Zheleva, L Getoor. [2008] Preserving The Privacy Of Sensitive Relationships In Graph Data, Proc. First Acm Sigkdd Int'l Conf. Privacy, Security, And Trust In Kdd, Pp. 153-171.
[3] S Nilizadeh, A Kapadia, YY Ahn.[2014] Community-enhanced de-anonymization of online social networks, in Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, ser. CCS '14. New York, NY, USA: ACM, pp. 537– 548.
[4] A Narayanan, V Shmatikov. [2009] 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, pp. 173–187.
[5] B Zhou, J Pei, W Luk. [2008] A brief survey on anonymization techniques for privacy preserving publishing of social network data,SIGKDD Explor. Newsl., 10(2): 12–22
[6] Mislove, B Viswanath, KP Gummadi, P Druschel. [2010] 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: ACMpp. 251–260.
[7] JK Jonghyuk Song, Jonghyuk Song. [2014] Inference attack on browsing history of twitter users using public click analytics and twitter metadata,IIEEE Transactions on Dependable and Secure Computing