A Mechanism to Preserve Private Data in Social Networking System

Lakshmi Vidyadharan
PG Scholar,
Department of CSE(PG),
Sri Ramakrishna Engineering College,
Coimbatore.

Karthigha M
Assistant Professor,
Department of CSE(PG),
Sri Ramakrishna Engineering College,
Coimbatore.

Mary Treesa Thomas
PG Scholar,
Department of CSE(PG),
Sri Ramakrishna Engineering College,
Coimbatore.

ABSTRACT

Social Networking System (SNS) like Facebook and Twitter have gained more popularity in this new era. It allows millions of individuals to create online profiles and share their personal information with vast networks of friend's. SNS allows third party extensions to access the users' information through Application Programming Interface (API). Since millions of users are using these sites it will lead to privacy problems and leakage of private information. This leakage happens without the knowledge of user, which leads to security problems like identity theft and phishing attack. Unknown user taking the information without our knowledge is called inference attack. This paper uses a permission based protection mechanism which limits the direct access of user data. Once an extension is certified by the user to access data from users' profile, then there is no more control on how it uses the data. Third party application can be built on trusted or un-trusted server. If it is an un-trusted server it will lead to inference attack and malicious user may use the information for unintended purposes and our data will be at risk. The main objective of this project is to enable the security access control scheme against inference attack.

Keywords

Social Networking System, Application Programming Interface and Online Social Networking.

1. INTRODUCTION

Network Security contains rules and regulations which are guided by network administrator who controls the authorization of access to data in a network. Users will be assigned with user ID and password to access the information. This will prevent unauthorized access, misuse, modification, or denial of a computer network and its resources. Network Security plays roles in different areas like business, government agencies, individuals, organizations, enterprises etc. It includes both private and public networks [8]. It protects the usability, reliability, integrity of network and also protects the network from various threats. Network Security can be accomplished through hardware and software. Software is constantly updated in order to protect from emerging threats. There are various components in network security and all these work together which helps to enhance the security. The different components are Anti-virus and anti-spyware, Firewall, Intrusion prevention systems (IPS), Virtual Private Networks (VPNs) which makes our network secured. Thus it secures the network, as well as it will check all operations being done.

1.1 Overview of the Project

Social networking system is a way to build social relations where people can share their interests, activities etc. It allows individuals to create a public profile and allow users to publish details about themselves and to connect to their friends. But some of the information is meant to be private. But it is possible to predict the private information from released data [5]. And if an unknown user or third-party which is running on untrusted server is taking the information without the users' knowledge, then it is called as inference attack. Here third-party access the users' information through Application Programming Interface (API) and permission based protection mechanism to protect the users' data. This will help us to understand whether the user is legitimate or not. However if the extension is authorized by the user, then there is no control on how it uses the data. This inference attack can also lead to many other serious issues like identity theft and phishing attacks. In this work, a comprehensive empirical study is done to assess the feasibility and accuracy of inference attacks and also an analytical framework for assessing the success rate of sample inference attacks.

1.2 Principles of Social Networking

Social Network Systems (SNSs) are one of the most popular application genres on the Internet. Facebook as the most favourite social network has around 900 million monthly active users (Facebook Newsroom) are today one of the most popular interactive medium to communicate, share, and disseminate a considerable amount of human life information. Daily and continuous communications imply the exchange of several types of content, including free text, image, audio, and video data. According to Facebook statistics1 average user creates 90 pieces of content each month, whereas more than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) are shared each month. The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data [9]. They are instrumental to provide an active support in complex and sophisticated tasks involved in OSN management, such as for instance access control or information filtering. Information filtering has been greatly explored for what concerns textual documents and, more recently, web content. However, the aim of the majority of these proposals is mainly to provide users a classification mechanism to avoid they are overwhelmed by useless data. In OSNs, information filtering can also be used for a different, more sensitive, purpose. This is due to the fact that in OSNs there is the possibility of posting or commenting other posts on particular public/private areas, called in general walls. Information filtering can therefore be used to give users the ability to automatically control 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, today OSNs provide very little support to prevent unwanted messages on user walls. For example, Facebook allows users to state who is allowed to insert messages in their walls (i.e., friends, friends of friends, or defined groups of friends). However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them. Providing this service is not only a matter of using previously defined web content mining techniques for a different application, rather it requires to design ad hoc classification strategies [15]. This is because wall messages are constituted by short text for which traditional classification methods have serious limitations since short texts do not provide sufficient word occurrences.

1.2.1 Access Control Mechanism for Third Party Application

Individuals can create their own profiles and share their personal information through these profiles. They can also subscribe to third-party applications to play games, edit photos etc. These developers host their application on their own servers. It will then interact with users and their personal information through Application Programming Interface (API). All the request sent by application is guarded a permission check. If the permission is granted by the user, then the requested data will be sent to the application.

1.2.2 Security Threat in the Online Social Networks

Third-party developers host their application on their own servers. These servers can be trusted or un-trusted servers. These applications can interact with the users and their profiles through Application Programming Interface (API). Every access via API is secured by permission check. If the permission is granted by the user then the application can access the user data. It has been reported that some of applications use the users' information other than the legitimate use. That is if the third-party application is hosted on the trusted server, then it is possible to control the misuse of user data. But if it is an un-trusted server then the user data will be at risk. A simple example is the selling of user information to marketing companies [10]. There is no guarantee that these applications will use user data in accordance with the purpose of applications. So if the user data is misused by the third-party application then it will be even more challenging problem.

1.3 Need For The Study

The Main objective of the proposed work is to enable the security access control scheme against interference attack. A naive interlocutor may argue that the above issue has already been addressed by the permission-based access control mechanism; in that third-party extensions cannot access user information without seeking the required permissions. If a user does not trust a third-party application, then the user shall not authorize it or use it. For instance, Facebook warns its users that when they subscribe to an application, they make their information accessible to a party which is neither known by Facebook nor by the user. This argument presumes that ordinary users have the necessary information and expertise to judge whether the applications they subscribe to are benign. In reality, most of the third-party applications are developed by developers who are not widely known to the user community. Not only that, the application is running on an entrusted server, meaning that there is no mechanism to monitor if the application is malicious. It is therefore not always possible for a user to assess if she can trust an application. It is our position in this proposed system that security-by-disclaimer is not a meaningful protection strategy. An interlocutor may also claim that SNS API inference attacks are but another minor privacy violation that does not warrant our attention. Here it disagrees for two reasons. First, analyzing the threats of any security or privacy concern must be accompanied by assessing the number of potential victims. If one develops a website with around 100 registered users, revealing their registration information means violating the privacy of only 100 users. However, when the number of potential victims reaches 50 million, they are facing a trouble with costly consequences. As mentioned earlier, popular Facebook applications may command a monthly active user count of 50 million. This implies that an inference attack with a meager success rate of 10% leads to privacy violations of 5 million victims. Second, SNS API inference attacks can be employed as a building block for conducting more dangerous security attacks. For instance, a well-known alternative authentication mechanism is to ask users a security question such as, "what is the name of your youngest sibling?", "who is your favorite author?", etc. Due to the nature of information that people upload to their SNS user profiles, answers to these security questions can usually be harvested systematically by launching inference attacks. The ability to answer a victim's security questions is the first step of identity theft [14]. Therefore, inference attacks could be an initial step in the launching of more dangerous attacks. Now, the question is who is best positioned to launch covert inference attacks. The answer is third-party extension developers. Gupta et al. (2013) recently showed that inference attacks can be employed to harvest the real interests of people and subsequently break mechanisms that use such personal information for user authentication.

2. PROPOSED SYSTEM

It has been reported (Steel and Fowler, 2010) that some applications use user information for purposes other than providing utility for users. If third-party applications were hosted on trusted servers, then it might be possible to control how they consume user data. Yet, third-party applications are hosted on untrusted servers. Delivering user data via the extension API to an untrusted server puts the data at risk. A simple example is the selling of user information to marketing companies. In short, there is no guarantee that third-party applications are benign, and that they use user data in accordance to the purpose of the applications[4]. While the above problem has to do with the misuse of legitimately accessible information, this work is instead about an even more challenging problem: i.e., through the extension API, malicious applications may obtain some private information for which they are not authorized.

2.1 Problem Objective

Here a permission based mechanism is used to access the user data. If the extension has been authorized once then there is no more control on how that extension uses the data. A malicious user will try to infer the private information which will lead to security issues[1]. 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 deep insight into the problem of Social Network Systems API inference attacks, and demonstrate the growing threat of such attacks to user privacy and also to evaluate the risk of SNS API inference attack[11].

2.2 Architecture Diagram

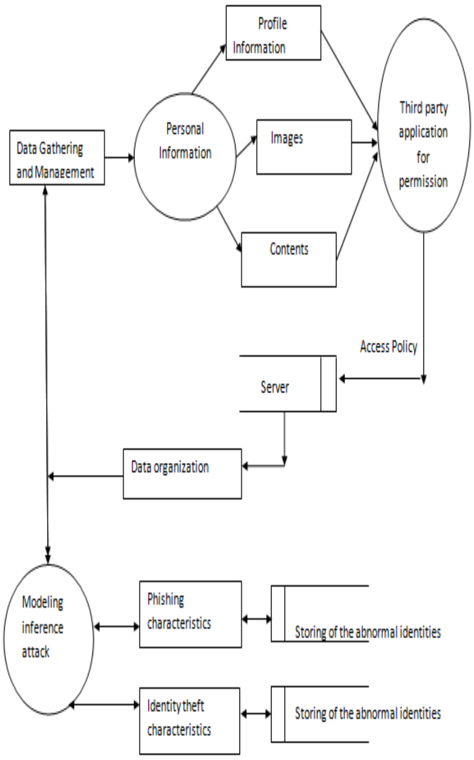

Profile Information
Third party application for permission
Data Gathering and Management
Personal Information
Images
Contents
Access Policy
Server
Data organization
Modeling inference attack
Phishing characteristics
Storing of the abnormal identities
Identity theft characteristics
Storing of the abnormal identities

Fig 2.1. Architecture of Proposed System

In an Online Social Networking various personal information will be given as shown in fig 2.1. This information includes profile information, images, and contents etc. This information will be taken by the third party application. For the third party application the user should give permission. If the permission is granted then the information will be stored on their corresponding server [13]. This server may be trusted or untrusted server. If it is a trusted server then the information is safe. Otherwise the information will be at risk. And this may lead to phishing attack and identity theft.

2.3 Advantages

- ❖ Correlation are weakened by randomization and its not been destroyed.
- ❖ Attribute reconstruction leads to the Preserving the sensitive data.
- ❖ Confidence level in the privacy in been measured with data perturbation.
- ❖ Information gain is negligible.
- ❖ Sanitation technique [3] is efficient against Sensitive data portioning.

3. RELATED WORK

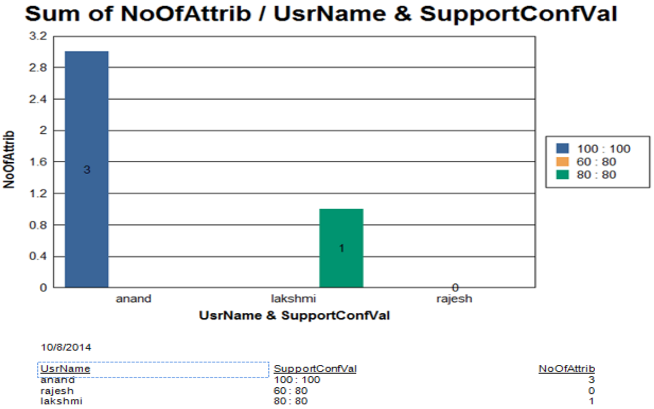
Raymond Heathery, Murat Kantarcioglu, and Bhavani Thuraisingham "Preventing Private Information sInference Attacks on Social Networks [9]" Online social networks, such as Facebook, are increasingly utilized by many people. These networks allow users to publish details about themselves and to connect to their friends. Some of the information revealed inside these networks is meant to be private. Yet it is possible to use learning algorithms on released data to predict private information. In this Literature, it explores how to launch inference attacks using released social networking data to predict private information. It then devises three possible sanitization techniques that could be used in various situations [9][3].

Gross.R, Acquisti.A, and Heinz. J.H,(2005) "Information Revelation and Privacy in Online Social Networks [6]" This paper deals with the study of information revelation in OSN and the privacy implications. Here it takes sample of around 4000 students who are using social networking sites frequently. After evaluating their information which they disclosed, it emphasize on probable attacks that can ensue on their information and also shows that only few users changes the highly leaky privacy preferences.

He J, Chu W, Liu Z (2006), "Inferring privacy information from social networks"[7] Currently, millions of individuals are sharing personal information and building social relations with others, through online social network sites. Recent research has shown that those personal information could compromise owners' privacy. In this work, it focus on the privacy of online social network users with missing personal information. It studies the problem of inferring those users' personal information via their social relations. It presents an iterative algorithm, by combining a Bayesian label classification method and discriminative social relation choosing, for inferring personal information.

4. EXPERIMENTAL ANALYSIS

In this Chapter, performance of the proposed system has been evaluated and proved its efficiency. Performance is calculated against the Not assisted versus assisted friend selection in privacy as a concern. Here it is represented as chart.



10/8/2014

UsrName	SupportConfVal	NoOfAttrib
anand	100 : 100	3
rajesh	60 : 80	0
lakshmi	80 : 80	1

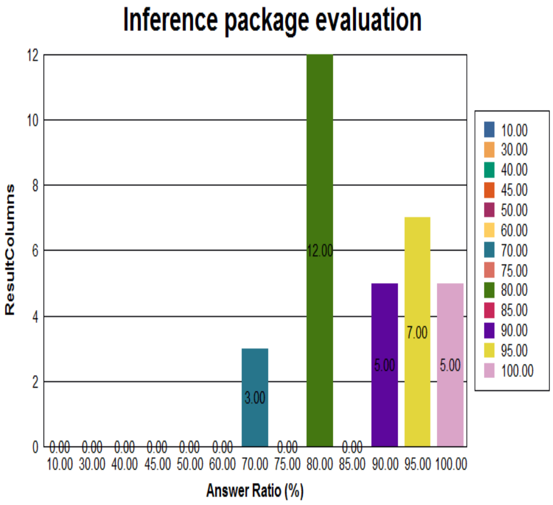
Figure 4.1 Chart Showing the Performance against the Not Assisted verses Assisted Friend Selection

4.1 Performance Analysis of the Proposed System against the Support Value with Number of Attributes.

Application of the direct usability is calculated with respect to the third party application extension. The first step in understanding a research problem is demonstrating its presence and relevance of the privacy of the user data against the inference attacks. To this end, two aspects of the problem should be clarified: (1) "how serious is the problem?" (Feasibility), and (2) "why is it a concern at all?" (Consequence). In this work, the paper took first step in understanding the problem of SNS API inference attacks.[11] This work is an attempt to give a clear view on the danger of SNS API inference attacks, and draw researchers' attention to this privacy concern. This paper conducts a comprehensive research study to show that SNS API inference attacks are not just a theoretical concern and implemented a OSN application with 10 inference algorithms and recruited more than 400 participants to reach 95% confidence level in our research experiment. It calculates SUPPORT and CONFIDENCE. The first factor for measuring the quality of an association rule is called support. It is obtained by dividing the number of records in S for which A and B are both true by the total number of records in S.

SUPPORT = (tuples containing both A and B) / (total number of tuples) (4.1)

CONFIDENCE= tuples containing both A and B) / (tuples containing A) (4.2)



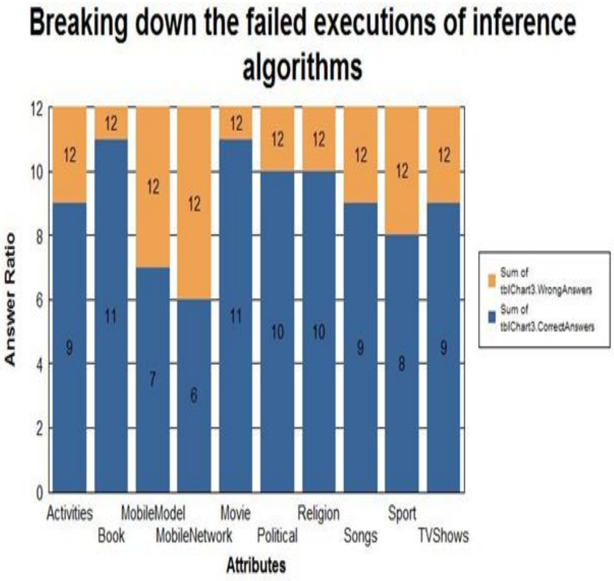
14-10-2014

AnswerRatio	ResultColumns
100.00	2
45.00	0
80.00	4
50.00	0
95.00	4
75.00	0
95.00	1
90.00	3
90.00	2
60.00	0
40.00	0
85.00	0
10.00	0
95.00	2
30.00	0
80.00	5
80.00	3
100.00	3
70.00	3

Figure 4.2: Chart showing Inference Package Evaluation

4.2 Performance of the Proposed System against Inference Resilience.

The result of the experiment revealed that not only SNS API inference attacks are feasible, but also they have considerably high success rates. Here it proposes an analytical framework to analyze the success rate of our inference algorithms for different classes of user profiles. This profile classification helps us understand the reasons behind high and low success rates of inference algorithms. If the support and confidence value is greater than minimum support threshold value, the OSN will provide the attributes which is requested from third party application.[11]. If the third party application user's answer is not fit to MST value it means inference attack indication is passed to the OSN and the user request from third party API is neglected.



Book	11	1
Movie	11	1
Political	10	2
Religion	10	2
Activities	9	3
Sport	8	4
Songs	9	3
TVShows	9	3
MobileModel	7	5
MobileNetwork	6	6

Figure 4.3 – Graph Showing Breaking Down the Failed Executions of Inference Algorithm

5. CONCLUSION AND FUTURE WORK

Solution developed and implemented a SNS with various issues related to private information leakage in social networks. Solution shows that using both friendship links and details together gives better predictability than details alone. In addition, solution explored the effect of removing details and links in preventing sensitive information leakage. In the process, Solution discovered situations in which collective inference does not improve on using a simple local classification method to identify nodes. When solution combine the results from the collective inference implications with the individual results, solution begin to see that removing details and friendship links together is the best way to reduce classifier accuracy[2]. This is probably infeasible in maintaining the use of social networks. However, solution also show that by removing only details, solution greatly reduce the accuracy of local classifiers, which give us the maximum accuracy that solution were able to achieve through any combination of classifiers [12]. Solution also assumed full use of the graph information when deciding which details to hide. Useful research could be done on how individuals with limited access to the network could pick which details to hide.

5.1 Future Enhancement

Similarly, future work scope could be conducted in identifying key nodes of the graph structure to see if removing or altering these nodes can decrease information leakage and also to show how to implement the right to erase the disclosed attributes when they are no longer necessary.

6. ACKNOWLEDGEMENTS

First and foremost I place this project work on the feet of GOD ALMIGHTY who is the power of strength in each step of progress towards the successful completion of my project. I am highly indebted to Dr.N.R.Alamelu, M.E., Ph.D. Principal, Sri Ramakrishna Engineering College, Coimbatore for providing the required facilities and support to carry out this project work. I extend my heartfelt thanks to Dr.A.Grace Selvarani, M.E., Ph.D. Professor and Head, Department of Computer Science and Engineering (PG), who had been a source of inspiration and for her timely guidance in the conduct of this project work. I am grateful to Mrs. KARTHIGHA, assistant professor, Department of Computer Science and Engineering (PG), for her able guidance and useful suggestions which helped me in completing the project work in time. Finally, I would like to express my heartfelt thanks to my beloved parents for their blessings, my friends for their help and wishes for the successful completion of this project.

7. REFERENCES

[1] 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.

[2] Ahmadinejad S, Anwar M, Fong P, "Inference attacks by third party extensions to social network systems" (2011) In: Proceedings of IEEE 9th International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 282-2877.

[3] Chandra D,Antony Rosewelt.L (2014) "Sanitations To Prevent Inference Attack On Social Network Data" , International Conference on Engineering Technology and Science-(ICETS'14) , Volume 3,pp.1265-1269

[4] Chen Li, Houtan Shirani-Mehr, Xiaochun Yang (2007) "Protecting Individual Information Against Inference Attacks in Data Publishing" vol. 4443, pp.422-433

[5] Facebook Newsroom. URL http://newsroom.fb.com/.

[6] Gross Ralph, Acquisti Alessandro,(2005)" Information revelation and privacy in online social networks", In: Proceedings of the 4th ACM Workshop on Privacy in the Electronic Society, WPES '05. ACM, pp. 71-80.

[7] He J, Chu W, Liu Z (2006), " Inferring privacy information from social networks" In: Intelligence and Security Informatics, Vol. 3975, pp. 154-165.

[8] Network Security, en.wikipedia.org/wiki/ network security

[9] Raymond Heathery, Murat Kantarcioglu, and Bhavani Thuraisingham(2013) "Preventing Private Information Inference Attacks on Social Networks" IEEE Transactions on Knowledge and Data Engineering, vol.25 pp. 1849- 1862

[10] Rajadurai .R, Nandini Priya .B, Swetha Sri .P (2014) "Overcoming Information Inference Attacks and Protecting Shared Data in OSN" Volume 4,pp. 414-417

[11] Seyed Hossein Ahmadinejad ,Philip W.L. Fong (2014)" Unintended disclosure of information: Inference attacks by third-party extensions to Social Network Systems" pp. 75-91.

[12] Thomas H. Hinke, Harry S. Delugach and Randall p.Wolf (1997) "Protecting databases from inference attacks" vol.16, pp.687-708.

[13] Luo W, Xie Q, Hengartner U. FaceCloak: an architecture for user privacy on social networking sites. In: Proceedings of the 12th IEEE International Conference on Computational Science and Engineering, Vol. 3 of CSE '09. IEEE Computer Society Press; 2009. pp. 26-33.

[14] Lucas MM, Borisov N. FlyByNight: mitigating the privacy risks of social networking. In: Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society, WPES '08. ACM; 2008.pp. 1-8.

[15] Zheleva E, Getoor L. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In: Proceedings of the 18th international conference on World wide web, WWW '09; 2009. pp. 531e40.