MEASURING PRIVACY IN ONLINE SOCIAL NETWORKS

Swathi Ananthula, Omar Abuzaghleh, Navya Bharathi Alla, Swetha Prabha Chaganti, Pragna chowdary kaja, Deepthi Mogilineedi

Department Of Computer Science and Engineering, University Of Bridgeport
Bridgeport, USA

ABSTRACT

Online Social Networking has gained tremendous popularity amongst the masses. It is usual for the users of Online Social Networks (OSNs) to share information with friends however they lose privacy. Privacy has become an important concern in online social networks. Users are unaware of the privacy risks involved when they share their sensitive information in the network.[1] One of the fundamental challenging issues is measurement of privacy .It is hard for social networking sites and users to make and adjust privacy settings to protect privacy without practical and effective way to quantify , measure and evaluate privacy. In this paper, we discussed Privacy Index (PIDX) which is used to measure a user's privacy exposure in a social network. We have also described and calculated the Privacy Quotient (PQ) i.e. a metric to measure the privacy of the user's profile using the naive approach. [2] The users should be aware of their privacy quotient and should know where they stand in the privacy measuring scale. At last we have proposed a model that will ensure privacy in the unstructured data. It will utilize the Item Response Theory model to measure the privacy leaks in the messages and text that is being posted by the users of the online social networking sites.

KEYWORDS

Online Social Security (OSN), Privacy Measurement, Privacy Index

1. INTRODUCTION

Information diffusion consists of a process in which a new thought or activity spreads through communication channels. Online Social Networks are the most used means for this nowadays *[3]*. Sociologists, marketers, and epidemiologists widely studied this area.

The OSN alludes to a network of independent IT consultants that utilize the network for an assortment of purposes, for example, information sharing, publicizing new opportunities, finding new companions. The OSN might likewise include companies that wish to make utilization of the services given by the consultants. Clearly, such companies ought to have a selective access to OSN resources. Hubs can likewise form smaller networks or groups. In the figure 1, the OSN is characterized as a directed labeled graph, where every hub resembles a network member and edges signify relationships between two separate members. Specifically, the initial hub of an edge indicates the member who established the relationship, while the terminal hub signifies the member who accepted to establish the relationship. Each edge is named by the sort of the relationship established and the corresponding trust level, representing how much the user that established the relationship trust the other user concerning that specific relationship. The OSN model interpreted in the figure comprises of three companies i.e. C_1, C_2 and C_3, whereas the remaining hubs represent agents. *[4]*

DOI : 10.5121/ijsptm.2015.4201

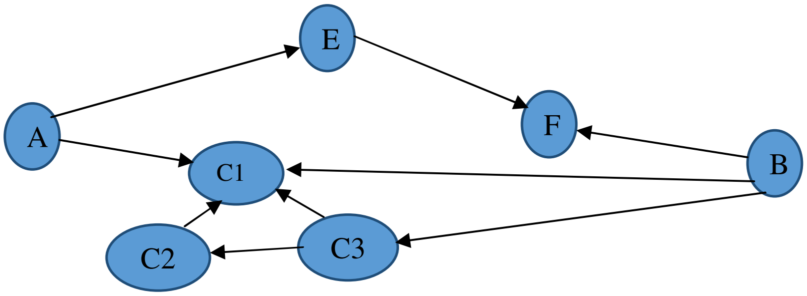

Figure 1: OSN model

The principle importance of an OSN is to make relationships with different users and accomplish such relationships for allocating resources of different nature. So, it is acknowledged that any access control model for OSNs ought to be relation-ship based*[5]*.

2. SECURITY ISSUES IN ONLINE SOCIAL NETWORKS

Clients give an astounding measure of individual data intentionally, and OSN administration suppliers store this data .Three primary people interact with each other in an OSN: the service provider, the users, and third-party applications.

2.1 Breaches from Service Providers:

OSN’s available client–server architecture intrinsically directs that clients must trust service provider to ensure all the individual data they've uploaded. However, service providers can clearly advantage from inspecting and imparting this data — for promoting purposes, for instance. Since service providers have the ability to utilize such data anyway they wish, researchers have raised genuine concerns and have endeavored to change this power imbalance. Researchers have proposed alternative OSN architectures as defenses. Scientists have proposed different option OSN architectures as resistances. *[6]* These proposition propose that end users ought to manage the fine-grained strategies with respect to who may see their data.*[7]*

Some Works were done by Service Providers in the area of protecting private information and enhancing security features with in social network sites. The most important solution is Locker. There are two observations under this solution. Firstly, we must use social relationships to describe access control policies. *[8]* Secondly, we must separate social networks from their content delivery and sharing. The idea is that Lockr proposals social network users' access control of their distribution data by hiding and mapping them into third-party storage. For example, images could be hidden in a storage server like Picasa6. The central idea with the Lockr extension is the necessity to be dependent on trusted third party storage for the hidden information.

2.2 Breaches from Other Users:

OSNs encourage interaction among friends. While satisfying this reason, service providers shield clients' security from unsubstantiated access. As an exchange off, all major OSNs let a client's friend get to the individual data the friend has transferred to his or her profile of course, while blocking other users from doing as such. Here, the approach of "friends" in an OSN is just a social connection that the two clients have consented to build in that OSN, paying little respect to the real disconnected from the net relationship. This disparity gives a potential channel to taking individual data by become a close acquaintance with clients in OSNs.For instance, 75,000 out of 250,000 irregular Facebook clients reached utilizing a programmed script acknowledged the script's appeal to turn into a Facebook friend. *[9]* Leyla Bilge and her associates have displayed two more-complex attacks.6 The first assault is called same-site profile cloning. An attacker copies a client's profile in the same OSN and uses the duplication to convey companion solicitations to the client's friend. Accepting the appeal has originated from a recognizable individual, the unalerted companions can acknowledge it and in this manner uncover their own data to the aggressor. The second assault is cross-site profile cloning. The aggressor identifies a client from OSN An, alongside this present client's companion list. The assailant then copies the profile to OSN B, where the client hasn't yet joined, and sends companion asks for on OSN B to the target's companions who have likewise enlisted on OSN B. Cross-site profile cloning is conceivably a larger number of hazardous than same-site cloning in light of the fact that its less inclined to stimulate suspicion. *[10][11]* At present, no guard can secure counter to such attacks. On the other hand, Leyla Bilge and her partners recommend expanding clients' readiness concerning their acknowledgement of companion requests.6 Also, enhancing the quality of Captcha can help avoid huge scale profile-cloning assaults utilizing robotized scripts.

2.3 Breaches from Third-Party Applications:

As OSNs grow their administrations, third-party applications are prospering in light of client requests for extra functionalities. Despite the fact that these applications live on the OSN stage, an outsider creates them, so they're basically untrusted. Also, clients must allow the application access to their own information before they can introduce those applications, in light of the fact that such get to is fundamental for a few applications to perform their usefulness. For instance, a horoscope application must know the client's birthday. Shockingly, neither the service provider nor the clients know precisely which bit of data is genuinely important for the applications. Thus, they must trust the applications to effectively proclaim the data they require. What's more, the component to screen how the applications control the individual data is missing. *[12]*This invites the applications to abuse that data.

3. RELATED WORK

3.1 Social Network Privacy Measurement Techniques

3.1.1 Privacy Quotient:

The unstructured data pose a problem for privacy score evaluation. The focus here is to evaluate a user’s privacy risks in exchanging, sharing, publishing, and disclosing unstructured data – namely, text messages.

A text message may contain sensitive information about user. The message is first checked for any sensitive information such as the user’s phone numbers, address, email, or location. The message is then classified as sensitive or non-sensitive by means of a naïve binary classifier. *[13]*.Each sensitive part of the message is treated as an “item” that has some sensitivity.

If a particular user j has shared information about the profile item i, then R(i,j)=1 i.e the information i is made public by the user j. If a particular user j has not shared information about the profile item i, then R(i,j)=0 i.e the information i is made private by the user j. Privacy quotient can be measured on two parameters i.e the sensitivity of the information and the visibility of the information. *[14]*

A.Calculation Of sensitivity

Sensitivity is the property of an information which makes it private. As sensitivity increases, privacy risks involved in sharing the item also increases. Hiding of such kind of information makes the user more private. Sensitivity (βi) of an item i can be calculated using the formula

$$\beta i = \frac{N - |Ri|}{N}$$

where |Ri| = j R(i,j) i.e the summation of all the cells of the column of profile item i where it has been made public. Figure 7 explains the same. On the basis of the data collected we have calculated the sensitivity of the profile items. The following table illustrates the same. *[15]*

B. Calculation of visibility

Visibility is the property of information that captures the popularity of an item in the network. The wider is the spread of information the more visible it is. V(i,j) i.e the visibility of a profile item i by the user j is calculated as;

V(i,j) = Pr[R(i,j)=1]X1+ Pr[R(i,j)=0] X 0;
V(i,j) = Pr[R(i,j)=1] + 0;
V(i,j) = Pr[R(i,j)=1];
where Pr[R(i,j)=1] = Probability that the value of R(i,j)=1;
and Pr[R(i,j)=0] = Probability that the value of R(i,j)=0;

C. Calculation of privacy quotient

If βi is the sensitivity of the profile item i and V(i,j) is the visibility of the profile item i for a user j. PQ(i,j) is the privacy quotient for a profile item i for user j and is calculated as; *[16]*

$$PQ(i, j) = \beta i * V(i, j)$$

To calculate the overall privacy quotient of the user j

$$PQ(j) = \sum PQ(i, j) = \sum \beta i * V(i, j)$$

where the range of items i.e. i varies from $1 \leq i \leq n$.

3.1.2 Privacy Armor - The Proposed Model To Ensure Privacy In Unstructured Data

We now propose a model that will measure privacy in the unstructured data in OSNs. The status updates, tweets and posts are all unstructured in nature. To calculate the percentage of privacy leaks in such data sets we have proposed Privacy Armor- a model that will warn the users if they are intentionally or unintentionally sharing some sensitive content online. *[17]*

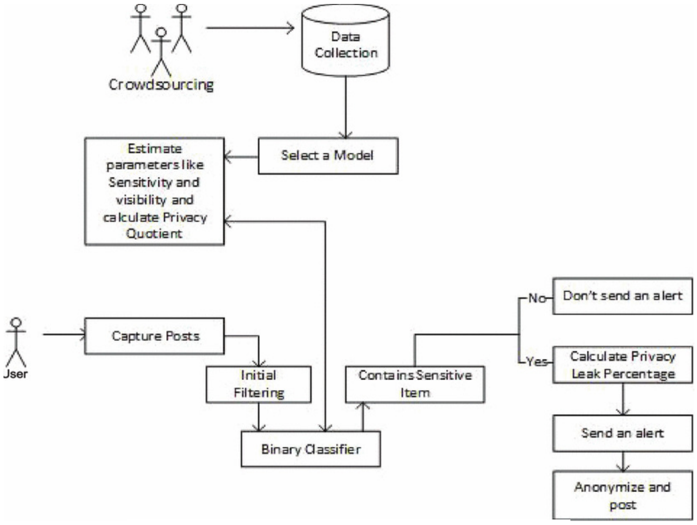

Figure 2: Proposed model of Privacy Armor

3.1.2.1 Crowdsourcing and Data Collection:

Initially using the crowdsourcing method we will gather the information about the items being shared on the user's profile. If they have willingly shared the data we will take it as 1 otherwise we will consider it as private entry and mark it as 0. The resultant will be a N X n dichotomous response matrix. Where N will be the number of users and n will be the number of profile items. *[18]*

3.1.2.2 Selecting a Model and calculating the Privacy Quotient

The advantage of naive approach is that it is a fairly simple approach and one can follow it easily as it has more practical implications. The disadvantage is that the sensitivity values obtained are significantly biased by the user population. If the users by nature are introverts and do not like to share a lot of information then the estimated sensitivity will be high, on the other hand if the users of the group are extrovert then the sensitivity will be low *[9]*. The real world data is too messy to fit the data effectively hence Liu et al have calculated the privacy scores by choosing the Item Response Theory model. To measure some trait of a person, there has to be a measurement scale *[16]*. An assumption that is made here is that every individual has some attitude, i.e., either the individual is an extrovert or an introvert. So users will have some attitude score that will place them somewhere on the attitude scale. This is denoted by θ. The probability that the jth individual having an attitude of θ will share their sensitive content i is denoted by P(θij). If we plot a graph with θ on the x axis and P(θij) on the y axis. It will a smooth S shaped curve that is called as the Item Characteristic Curve. *[19]* This curve has two properties, one is the sensitivity that is denoted by β and the other is the discrimination constant denoted by α. Privacy quotient can be calculated as stated in equation 4. Using the item response theory model the V(i,j) is calculated as

$$\mathrm{PR}(\theta ij\ = 1) = \frac{1}{1\ +\ e\hat{}(\alpha i(\theta j - \beta i))}$$

where βi is the sensitivity of the ith profile item, αi is the discrimination constant of the ith profile item, θj is the ability of the jth user. The calculated values of the parameters like the sensitivity, visibility are highly intuitive. This computation can also be parallelized using the Map Reduce technique, which can thereby increase the performance of the algorithm as well. After calculating the sensitivity and visibility. We can compute the privacy quotient of each of the users using the equation 4. Sharing of messages in the form of status updates, tweets etc is very common now-a-days. Such information may contain some sensitive information about the user *[17] [21].* Some users intentionally share it whereas some users are not aware of the privacy risks that follows. Privacy armor will warn such users and send them an alert showing the privacy leakage percentage. In Figure 10 the message posted by the user is first analyzed by the privacy armor to check for any sensitive information such as their phone numbers, email, address, location etc. By making use of a binary classifier the posts are either classified as sensitive or not sensitive. Also a percentage of privacy leakage is shown to the users. Privacy leakage is calculated as

$$\vartheta = \frac{\sigma}{\beta} * 100$$

Where σ is kσi here k are number of sensitive items in the post and σi is the sensitivity of the i th profile item. β is the total sensitivity of all the n items. ϑ is the percentage of privacy leakage. For eg: If the user shares "Having lunch with Congress supporters". *[22]* Here the user is sharing the political view. A certain amount of privacy leak is associated with this post which can be calculated using equation 6. As calculated by the naive approach the sensitivity of political views is .6833 and the overall sensitivity is 4.183. ϑ = (.6833/4.183)*100; ϑ = 16.33 %; Privacy leakage associated with this post is 16.33 %.

People with low privacy quotient, are likely to share the information without considering the privacy risk. Sharing information with such users is often risky and will have a high percentage of privacy leaks.

3.1.3 Privacy Index (PIDX):

Privacy Index (PIDX) is used to describe an entity's privacy exposure factor based on known attributes in actor model. We expand PIDX to social network model to measure an actor's privacy exposure to another. *[23][24]*

Privacy Index PIDX(i,j) is used to describe actor Aj 's privacy exposure to Ai based on Aj *'s* visible attributes to Ai . High PIDX value indicates high exposure of privacy. Privacy Index $PIDX$ is between 0 and 100. PIDX value can be used for privacy monitoring and risk control.*[23][25]*

PIDX is defined as the ratio of the sum of the privacy impact factors of the published items, set K, to the sum of the privacy impact factor of all the items, set I.

The privacy index PIDX computation for the user is the same as the computation of a message's privacy leakage of Privacy Quotient because sensitivity of an item i is Si = βi. *[25]*

3.2 SONET model:

An online social network may attract millions of clients. These clients are linked together through ties with friend. Every client can be described by client profile, privacy settings, and a friend list. A profile comprises of personal information of a client. Privacy settings describe how clients need to convey their own information. A friend list includes a group of individuals who are connected together. The friend list can be further classified as different groups, for example, friends, friends of friends, or public, etc. Privacy settings can thus be characterized according to the groups. *[26]* SONET model provides an effective and practical way to model privacy in social networks. Figure 1 shows a mapping between a social network and the SONET model. Only two users are demonstrated in the figure. *[27]*

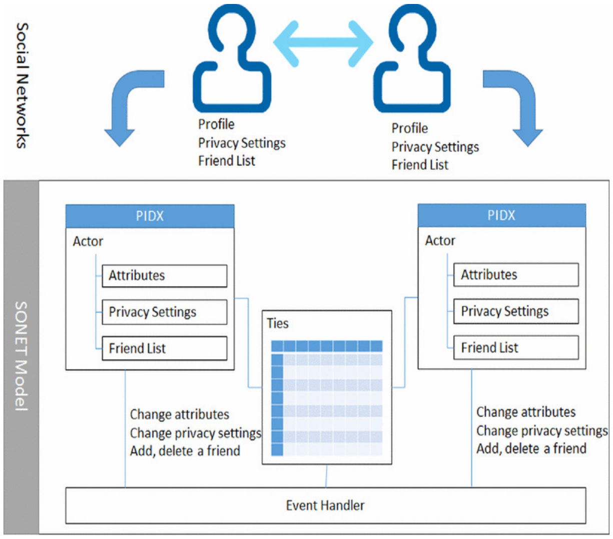

Figure 3: SONET Model for Social Networks

In SONET model, clients are represented as actors. Profile is depicted by their attribute list. The attributes are further extended with hidden data and virtual attributes. Friend list is depicted by a degrees of separation function $h(Ai\ ,\ Aj)$. Privacy settings are depicted by a attribute visibility function (, d) which gives back a numeric value to show that if a specific attribute is visible to another actor. PIDX can accordingly be assessed to reflect a client's privacy presentation. Further, SONET model also supports events, for example, attribute value changes, privacy settings update, and friend add/delete. SONET model can be utilized to stimulate privacy changes in an social network and assess privacy affect in case clients accounts are compromised.

3.3 PrivAware:

PrivAware is a tool to detect and report unintended information loss in social networks. PrivAware calculated privacy score based on the total number of attributes visible to the third party applications to the total number of attributes of a user. It does not consider sensitivity of an attribute.

3.4Privometer

Privometer is used to measure the amount of sensitive information leakage in a user’s profile. The leakage is indicated by a numerical value. The model of Privometer also considers substantially more information that a potentially malicious application installed in the user’s friend realm can access. *[28]*

4. CONCLUSION

Privacy measurement is a perplexing concern in social networks. In this paper, we cover SONET model to support privacy measurement in social networks. *[28]* We propose to use PIDX(x, y) to measure actor Ay ’s privacy exposure to Ax . We Measure Privacy by taking sensitivity and visibility attributes in to consideration. SONET model gives experimental and efficient way for online users and social networking sites to measure privacy.

REFERENCES

[1] R. Gross and A. Acquisti, “Information revelation and privacy in online social networks,” in Proceedings of the 2005 ACM workshop on Privacy in the electronic society. ACM, 2005, pp. 71–80.

[2] L. Backstrom, C. Dwork, and J. Kleinberg, “Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography,” in Proceedings of the 16th international conference on World Wide Web. ACM, 2007, pp. 181–190.

[3] J. DeCew, “Privacy,” in The Stanford Encyclopedia of Philosophy, E. N. Zalta, Ed., 2012. [4] Y. Altshuler, Y. Elovici, N. Aharony, and A. Pentland, “Security and privacy in social networks.” Springer, 2013.

[4] M. Huber, M. Mulazzani, E. Weippl, G. Kitzler, and S. Goluch, “Exploiting social networking sites for spam,” in Proceedings of the 17th ACM conference on Computer and communications security. ACM, 2010, pp. 693–695.

[5] P. Gundecha and H. Liu, “Mining social media: A brief introduction.”

[6] B. Fung, K. Wang, R. Chen, and P. S. Yu, “Privacy-preserving data publishing: A survey of recent developments,” ACM Computing Surveys (CSUR), vol. 42, no. 4, p. 14, 2010.

[7] R. Agrawal and R. Srikant, “Privacy-preserving data mining,” in ACM Sigmod Record, vol. 29, no. 2. ACM, 2000, pp. 439–450.

[8] K. Liu and E. Terzi, “A framework for computing the privacy scores of users in online social networks,” in Data Mining, 2009. ICDM’09. Ninth IEEE International Conference on. IEEE, 2009, pp. 288–297.

[9] L. Fang and K. LeFevre, “Privacy wizards for social networking sites,” in Proceedings of the 19th international conference on World wide web. ACM, 2010, pp. 351–360.

[10] A. Braunstein, L. Granka, and J. Staddon, “Indirect content privacy surveys: measuring privacy without asking about it,” in Proceedings of the Seventh Symposium on Usable Privacy and Security. ACM, 2011, p. 15.

[11] S. Guo and K. Chen, “Mining privacy settings to find optimal privacyutility tradeoffs for social network services,” in Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom). IEEE, 2012, pp. 656–665.

[12] J. L. Becker and H. Chen, “Measuring privacy risk in online social networks,” Ph.D. dissertation, University of California, Davis, 2009.

[13] J. Anderson, “Privacy engineering for social networks,” 2013.

[14] H. R. Lipford, A. Besmer, and J. Watson, “Understanding privacy settings in facebook with an audience view.” UPSEC, vol. 8, pp. 1– 8, 2008.

[15] F. Drasgow and C. L. Hulin, “Item response theory,” Handbook of industrial and organizational psychology, vol. 1, pp. 577–636, 1990.

[16] H. Mao, X. Shuai, and A. Kapadia, “Loose tweets: an analysis of privacy leaks on twitter,” in Proceedings of the 10th annual ACM workshop on Privacy in the electronic society. ACM, 2011, pp. 1–12.

[17] J. Becker, “Measuring Privacy Risk in Online Social Networks,” Design, vol. 2, p. 8, 2009.

[18] E. M. Maximilien, T. Grandison, T. Sun, D. Richardson, S. Guo, and K. Liu, “Privacy-as-a-Service : Models , algorithms , and results on the facebook platform,” in Web 2.0 Security and privacy workshop, 2009.

[19] K. U. N. Liu, “A Framework for Computing the Privacy Scores of Users in Online Social Networks,” Knowl. Discov. Data, vol. 5, no. 1, pp. 1–30, 2010.

[20] N. Talukder, M. Ouzzani, A. K. Elmagarmid, H. Elmeleegy, and M. Yakout, Privometer: Privacy protection in social networks, vol. 1, no. 2. VLDB Endowment, 2010, pp. 141–150.

[21] E. M. Maximilien, T. Grandison, T. Sun, D. Richardson, S. Guo, and K. Liu, “Privacy-as-a-service: Models, algorithms, and results on the facebook platform,” in Proceedings of Web, 2009, vol. 2.

[22] C. Akcora, B. Carminati, and E. Ferrari, “Privacy in Social Networks: How Risky is Your Social Graph?,” in 2012 IEEE 28th International Conference on Data Engineering, 2012, pp. 9–19.

[23] J. Bonneau and S. Priebusch, “The Privacy Jungle : On the Market for Data Protection in Social Networks,” in The Eighth Workshop on the Economics of Information Security, 2009, pp. 1–45.

[24] R. N. Kumar and Y. Wang, “SONET: A SOcial NETwork Model for Privacy Monitoring and Ranking,” in The 2nd International Workshop on Network Forensics, Security and Privacy, 2013.

[25] Y. Wang and R. N. Kumar, “Privacy Measurement for Social Network Actor Model,” in The 5th ASE/IEEE International Conference on Information Privacy, Security, Risk and Trust, 2013.

[26] M. S. Ackerman, L. F. Cranor, and J. Reagle, “Privacy in ecommerce: examining user scenarios and privacy preferences,” in Proceedings of the 1st ACM conference on Electronic commerce, 1999, vol. 99, no. 1998, pp. 1–8.

[27] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and analysis of online social networks,” Proc. 7th ACM SIGCOMM Conf. Internet Meas. IMC 07, vol. 40, no. 6, p. 29, 2007.

[28] L. Sweeney, “Uniqueness of simple demographics in the U. S. population,” in Data privacy Lab white paper series LIDAP-WP4, 2000.