2016 International Conference on Internet of Things and Applications (IOTA)
Maharashtra Institute of Technology, Pune, India 22 Jan - 24 Jan, 2016

Personalization in User Profiling:

Privacy and Security Issues

Bhushan S. Atote
Computer Department
MAEER's MIT
Kothrud, Pune
bhushanatote2408@gmail.com

Saniya Zahoor
Computer Department
MAEER's MIT
Kothrud, Pune
saniya.zahoor@yahoo.com

Bharat Dangra
Computer Department
MAEER's MIT
Kothrud, Pune
bharatdangra@gmail.com

Dr. Mangesh Bedekar
Computer Department
MAEER's MIT
Kothrud, Pune
mangesh.bedekar@gmail.com

Abstract— Due to the sheer volume of data available on the internet, the problem of information overload is an ever increasing one. Personalization systems based on user profiles tend to assist the user in his day-to-day tasks by recommending 'relevant' content to the user based on his previous web usage and navigation patterns. These user profiles hold more and more information about the user, his likes, dislikes as it learns about the user over time. Privacy and security issues of the 'user profiles' becomes far the more important as the user profiles actually reflect the user itself albeit in a pseudo form.

Keywords— User profiles, Personalization, Privacy and security issues, Private cloud, Anonymization.

I. INTRODUCTION

The Internet has made information available to humankind in a quick, easy, publicly accessible manner which is within reach of one and all. Today Internet is the first resource any user turns to when he needs any form of information. A multitude of web pages exist on any topic and the number grows with each passing day. The task of finding relevant information from the corresponding web pages is a tedious task. Search engines are available for the same but the problem of ever increasing data reduces the efficiency of the search results performed by a user. Even if the user manages to find a relevant page corresponding to his search query, retaining and remembering the webpage, or storing it efficiently for future references is another task not very well achieved. Identifying the relevance of a web page to a user thus becomes a very challenging task.

User Profiling provides the means to obtain the accurate information about the user's interests and represent it with minimal user intervention. It helps to address the information overload problem disambiguates web search terms, deliver personalized web search results and make the web more relevant and personal. Also, Dynamism of user's behavior can be met and it helps in managing Privacy issues as well. But there is major drawback of privacy issues. Big data techniques provides excellent tools and techniques for more accurate User Profiling. However, there is an issue of privacy in User Profiling. Here we need to develop techniques that can collect information for User Profiles at one end, while respecting the privacy of the users at the other end. This will in turn increase the level of confidence of users toward such systems. A lot of research is currently in progress in this area. Preserving the private information about the user is a real challenge. In this paper we elaborate on various privacy and security issues with regards to the 'user profiles' and suggests some solutions for the same.

II. LITERATURE SURVEY

User profiling is very beneficial to users as it has many advantages to offer when compared with the day to day search engines. Privacy is a drawback in User Profiling if it is not done correctly. Various privacy concerns and suggested solutions are discussed below:

A. PRIVACY PROBLEMS

A lot of research has been done in the field of preserving the user privacy. Some are discussed below:

[1] have discussed User Profiling privacy challenges. User Profiling is the process of collecting information about a user so as to know him very well and recommend him with more accurate and relevant information. The information contained in User Profiles like geographical location, professional background, interests, preferences, etc. is personal and there are privacy issues related to it. One of the major issues is the selling of personal information in User Profiles to third parties for profit, for commercial or even malicious purposes. Similarly there are other issues discussed in the paper like utilization of User Profile data for targeted advertising and so on. [2] have focused on privacy issues of Personalization. They have focused on personalized content in a social networking system that can reveal private information directly to others. They have also focused on issues like personalizing the content according to the physical location of the user that can reveal his location to others, which will again interfere with the privacy issues of users. [3] proposed a new class of statistical deanonymisation attacks against high-dimensional micro-data successfully and have identified the Netflix records of known users, uncovering their potentially sensitive information using the Internet Movie Database as the source of background knowledge. [4] discussed the privacy issues of social networking sites like Facebook. They have studied two of Facebook's features - News Feed and applications. News Feed involves not just questions of privacy, but also of program interface and of the meaning of ''friendship'' online. Second, Applications shares information about the user as well as his friends' information with third parties, which is again a security issue. [5] focused on cloud

computing security concerns, especially data security and privacy protection issues.

B. SOLUTIONS

Privacy issues can be handled well if we apply the various security related techniques in User Profiling. Some of them are discussed below:

[6] have introduced well-known approach of k-anonymity in privacy preservation of published data. It focuses on the separation of information into sensitive attributes and quasi-identifiers. Sensitive attributes are those that contain personal privacy information like disease, salary, etc. while Quasi-identifiers determine the identity of the individuals referred to by a record like gender, birthday, etc.

[7, 8] have introduced the concept of Pseudonymous personalization as a solution for privacy issues in user profiling. They have introduced the approach that enables the system to keep the track of the same pseudonym across different sessions and provide personalized services without knowing the true identity of the pseudonym, thus, ensuring full privacy. [9, 10, 11, 12, 13, 14] have focused on client-based personalization that stores the users' information at the client side i.e.; users' computers, mobile phones, etc... Thus, the physical security of user's device is the only concern about this solution which is in the hands of user itself.[15] have focused on privacy solutions in recommendation systems and have come up with a number of techniques like aggregation, distribution, perturbation and obfuscation that protects user's privacy in recommender systems. [16] suggested putting suitability into user modeling and personalized systems that help in privacy protection.

[17, 18, 19, 20] have focused on privacy-preserving location tracking to enhance location privacy and hide the location of users, which in turn will hide device's visited locations from others. Thus, security will be ensured to the private data about the user.

III. OUR PROPOSED SYSTEM WITH ISSUES AND SOLUTIONS

We are working on User Profiling of a user on different devices and all this will be synchronized via Cloud. User may use, Laptop, Desktop, Mobile and so on as shown in the figure 1 below:

User	U1	U2	U3	U4
Devices	Mobile, Laptop	Laptop, Mobile, Desktop	…	…
Profiles	P11, P12	P21, P22, P23	…	…

Fig. 1. Association between User, Device and Part Profiles

Thus, Single User's Profiling, has to be done on multiple devices that he uses. For that we need to model a single User by many Part Profiles that best represent the User.

For implementation, we are going to use Mozilla Firefox and Greasemonkey as the frontend. Mozilla Firefox is a browser where a user will perform his actions and Greasemonkey is an add-on on the Mozilla Firefox where we will write our scripts to capture those actions of the user. JavaScript is used to write the scripts. For the Backend, we are going to use XAMPP to store all the data of the user and PHP language which is a scripting language.

Now once all this is implemented in all the devices of the user, we will get many part profiles of a single user. His purpose of use and behavior will be different on different devices [hence different profiles]. Aim is to get the combination of his Part Profiles to get the best of the profile for him. There is a need to synchronize his behaviors [on different devices] and get the best combination of profile behaviors for him. It's the same user and these are his tasks which though disjoint are related and relevant. The more the user accesses different devices, the more detailed behaviors we will and synchronizing them will lead to better and richer profiles and hence more relevant information can be given to a user. Even if one Device fails [is changed] the user still can get the relevant information from other device.

However, there are privacy concerns in our proposed system that needs to be taken care of. We consider the User Profile to be composed of three layers – Public Layer (Outer Layer), Private Layer (Intermediate Layer) and Personal Layer (Inner Layer) as shown in figure 2.

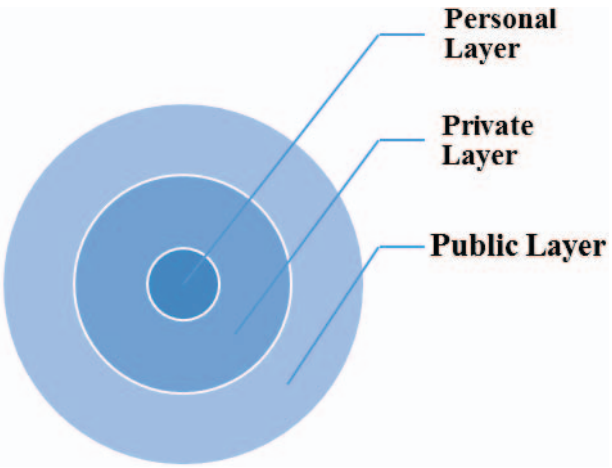

Fig. 2. Layered Structure of User Profile

These are discussed as follows:

- PUBLIC LAYER: This is the outermost layer. This layer has Minimum privacy concerns. Example would be your name, gender of the user.
- PRIVATE LAYER: This is the intermediate layer. This layer has privacy concerns. Example would be telephone number, hometown, and blood group of the user.
- PERSONAL LAYER: This is the innermost layer. This layer has high privacy concerns. Example would be sending a mail to your 'friend'.

We need to protect the private and personal layer. For that, we will categorize the information about the user into above three categories and employ different means and degrees of security related techniques to the three layers. This will help us to protect the various security attacks on the information in

User Profile. We can use various mechanisms that we will help us to ensure privacy in our system. These are:

- ENCRYPTION: All the data will be stored in some form not readable to the others.
- PRIVATE CLOUD: Since we need a cloud to synchronize the part profiles and by using a public cloud can introduce a security breech in our system. Therefore, there is a need of private cloud.
- PART PROFILES AT DIFFERENT PLACES: We can store the part profiles at different places. Even if the part profile is seen by someone, he might see only one part profile, the rest will be secured.
- CLIENT SIDE PRIVACY: Since the data will be stored at the client side first, then it will be synchronized over the cloud. So the personal security of the device also needs to be taken care of by the user.

These techniques will look after the privacy concerns of our proposed system. We are proposing some approaches of applying these techniques in our system. In the first approach, we store the entire user's information at client side and then categorize the information into public, private and personal information. Various degrees of encryption methods can then are applied to these information(s). This information content can be stored on a private cloud. Another variation is to store the three parts of information separately on private cloud. This will provide better security to the information in user profiles.

IV. CONCLUSION

User satisfaction is the ultimate aim of personalization. User Profiling is the process of collecting the user-information from the interaction done on the device, which is then used to recommend appropriate content to the users. Once the user's needs are established, techniques, such as "content filtering" and "collaborative filtering", are used to decide what content might be appropriate.

User profiling has many advantages, if done correctly, but there is a major concern of security in such systems. For that we can have techniques that need to be used to take care of security issue in such systems. For example, all the data about the user must be stored in an encrypted form so that it cannot be used by others. There are bulks of encryption techniques that can be used to ensure the security of data. Thus User Profiling is beneficial if done correctly without interfering in the personal details of the user.

Correct privacy and security measures need to be taken to ensure that the user's personal information in the profile is safe.

References

[1] Omar et. al., A Discussion of Privacy Challenges in User Profiling with Big Data Techniques: The EEXCESS Use Case. Presented at IEEE 2nd International Congress on Big Data, Santa Clara Marriott, CA, USA, June-July 2013.

[2] Eran Toch, Yang Wang, Lorrie Faith Cranor, Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems, Journal of User Modeling and User-Adapted Interaction, Volume 22, Issue 1-2, pp. 203-220, April 2012.

[3] Arvind Narayanan, Vitaly Shmatikov, Robust De-anonymization of Large Sparse Datasets (How To Break Anonymity of the Netflix Prize Dataset), IEEE Security and Privacy, Oakland 2008.

[4] Gordon Hull, Heather Richter Lipford, Celine Latulipe, Contextual Gaps: Privacy Issues on Facebook, Ethics and Information Technology Vol. 13, No. 4, pp. 289-302, 2011.

[5] Deyan Chen, Hong Zhao, Data Security and Privacy Protection Issues in Cloud Computing, International Conference on Computer Science and Electronics Engineering, 2012.

[6] M. Bilenko, M. Richardson, Predictive client-side profiles for personalized advertising, KDD'11, Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, August 21-24, San Diego, California, USA, Pp. 413-421, 2011.

[7] Cassel, Wolz, Client side personalization, Proceedings of the joint DELOS-NSF workshop on personalization and recommender systems in digital libraries, Dublin City University, Dublin 2001.

[8] Ceri, S., Dolog, P., Matera, M., Nejdl, W., Model-driven design of web applications with client-side adaptation. International Conference on web engineering, ICWE'04, vol. 3140, Springer, Munich, pp. 201–214, 2004.

[9] Coroama, V., Langheinrich,M., Personalized vehicle insurance rates—a case for client-side personalization in ubiquitous computing. Ubiquitous Comput. Workshop Privacy Enhanced Personal. CHI' 06(22), pp. 56–59, 2006.

[10] Hitchens, M., Kay, J., Kummerfeld, B., Brar, A. Secure identity management for pseudo-anonymous service access. In: Hutter, D., Ullmann, M. (eds.), Security in pervasive computing: second international conference, Boppard, pp. 48–55, 2005.

[11] Arlein R.M., Jai B., Jakobsson M., Monrose F., Reiter M.K.: Privacy-preserving global customization, In 2nd ACM conference on electronic commerce, ACM Press, Minneapolis, pp. 176–184, 2000.

[12] Mulligan, D., Schwartz, A.: Your place or mine?: privacy concerns and solutions for server and client-side storage of personal information. In: Proceedings of the tenth conference on computers, Freedom and privacy: challenging the assumptions, ACM Press, Toronto, pp. 81–84, 2000.

[13] Gerber, S., Fry, M., Kay, J., Kummerfeld, B., Pink, G., Wasinger, R. Personis J: mobile, Client-Side user modelling. In: International conference on user modeling, adaptation, and personalization, lecture notes in computer science, vol. 6075, Springer, Berlin, pp. 111–122, 2010.

[14] Schafer, J., Frankowski, D., Herlocker, J., Sen, S., Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.), The Adaptive Web, Springer-Verlag, Berlin, pp. 291–324. 2007.

[15] Kay, J., Scrutable adaptation: because we can and must. In: Adaptive hypermedia and adaptive web-based systems, Springer, Berlin, pp. 11–19, 2006.

[16] Beresford, A.R., Stajano, F.: Location privacy in pervasive computing. IEEE Pervasive Computing, 2(1), pp. 46–55, 2003.

[17] Gruteser, M., Liu, X.: Protecting privacy, in continuous location-tracking applications. Security Privacy, IEEE. 2(2), pp. 28–34, 2004.

[18] Hoh, B., Gruteser, M.: Protecting location privacy through path confusion, In: Security and privacy for emerging areas in communications networks, 2005, SecureComm 2005, First international conference on security and privacy for emerging areas in communications networks, IEEE Computer Society, Washington, pp. 194–205, 2005.

[19] Ristenpart, T., Maganis, G., Krishnamurthy, A., Kohno, T.: Privacy-preserving location tracking of lost or stolen devices: cryptographic techniques and replacing trusted third parties with DHTs. In: Proceedings of the 17th conference on security symposium, USENIX Association, San Jose, pp. 275–290, 2008.

[20] Gedik, B., Liu, L., Protecting location privacy with personalized k-anonymity: architecture and algorithms. IEEE Transactions on Mobile Computing 7(1), pp. 1–18, 2008.