TPP: Tradeoff between Personalization and Privacy

Ubaid Ur Rehman[1] and Sungyoung Lee[2]

Kyung Hee University (Global Campus), Yongin-si, Republic of Korea,
[1]ubaid.rehman@khu.ac.kr,[2]sylee@oslab.khu.ac.kr

Abstract. Modern technology relies on personalization due to its appealing services. It suggests the most relevant information to the users. Beside it several benefits, there may be some privacy leakage due to the personalization. As it analyzes and collects the user behavior's data, and generates a personalized decision. In this paper, we have considered the personalization aspect of recommendation, crowdsensing, and healthcare domains. We have identified the state-of-the-art research, specifically emphasizing on the personalization and privacy aspect. Also, we have conducted a survey, in order to identify the literacy of personalization and privacy. Moreover, we have discussed the attacks that exploit the vulnerability of personalization.

Keywords: Personalization, Crowdsensing, Recommendation, Healthcare, Privacy

1 Introduction

With the emergence of technology, online social networking platforms have become an important aspect of every individual. Several social media platforms are available, such as Facebook, Twitter, and Flickr, which can be used for networking, microblogging, and site tagging respectively. These platforms facilitate users in different degree of interaction, which includes resource sharing, chatting, online gaming, and other services. Due to the direct involvement of users, a large quantity of social information gets generated that provide assistance in many situations, such as user feedback regarding a social concern or product helps in retrieving an accurate information. In order to understand the preferences and behavior of a user, these quality social information are used by different technologies to provide a personalized suggestion.

Personalization helps the user to retrieve the information efficiently based on their preferences. The most common example is the recommender system, which generates a personalized recommendation based on the user browsing behavior. It has been applied in a variety of domains such as online multimedia, news, shopping, social media, and tourism. According to [24], 90% of marketer has deployed personalization strategies, due to the high impact on economy. Beside the recommender system, the crowdsensing and healthcare domains also use the personalization aspect to facilitate the users. Crowdsensing utilizes mobile phone

sensors to collect user data, instead of deploying a sensor network. The sensing is classified into personal and community, which monitor the individual and group of users respectively. Currently, most of the healthcare applications use this concept of crowdsensing to monitor patient vitals, behaviors, and emotion. Based on these constraints, the application recommends a diagnosis, treatment, or follow-up plan for the patient.
Despite the fact of several benefits, privacy is considered a secondary requirement in these domains. Most of the applications focused on efficient, reliable, and accurate decision making/recommendation. According to Jeong et al. [25], users are interested in personalization, but they are also curious about how the service provider handles their personal data. Therefore, in the case of a recommender system, a small hint may spoil a big surprise. Suppose, in a husband-wife relationship, the husband bought an expensive birthday gift for his wife using a shared online shopping account. The purpose was to surprise her, but when his wife login to the online shopping account, the system will provide a similar recommendation based on the recent purchase, which will spoil the surprise. Similarly, in crowdsensing and health domain, mobile phone's sensors and computation are used, where the data leakage regarding a specific diagnosis may lead to a serious consequence. Most of the users desired high-quality personalization that required more personal data, but the users are unwilling to share their personally identifiable information. Therefore, we have emphasized on the tradeoff between personalization and user privacy, targeting recommender system, crowdsensing, and healthcare domains. In this research study, we have focused on the following research questions (RQs):

RQ1: How personalization is deployed in recommender system, crowdsensing, and healthcare domains?
RQ2: How personalization will affect the privacy?
RQ3: What will be the tradeoff between personalization and privacy?

The rest of the paper is classified as section 2 describes the state-of-the-art techniques that use the personally identified information. Section 3 represents the survey result regarding the data privacy awareness in the community. The attacks that can occur due to personalization is discussed in section 4. Finally, section 5 will summarize the conclusion and future work.

2 Related Work

The personalization has been evolved and attracted great attention of the research community. We have covered the literature of recommender system, crowdsensing, and healthcare domains. Based on our literature review, we have selected the most relevant papers that cover the aspect of personalization.

2.1 Recommender System

Most of the personalized recommender systems, applications, engines, and frameworks have been proposed in the last few years. This mostly endorses the user to use their social network's account to avail the services. The goal is to collect and analyze the social network's activity and then generate a personalized recommendation accordingly. We have classified the recommendation systems into *a*) Everyday Items, *b*) Like-minded People, and *c*) Article Suggestion.

a) Everyday Items: In everyday items recommendation, the user is considered as an independent entity and suggested with a few random items, which leads to a cold start problem. In order to tackle the cold start problem, many solutions have been proposed [9]. A little interaction or social network activity analysis leads to an effective and personalized recommendation [10]. Similarly, Guy et al. in [12] described the relationship between the user and the items using the proposed ranking function.

b) Like-Minded People: The social network platforms compared the user profile contents and recommend based on the similarity index. Groh et al. [14] used the neighborhood information from the social network and analyzed statistically using a collaborative filtering method. A unified framework for user recommendation was proposed in [15]. In [13], the authors designed an approach that collects information from different sources and generates a recommendation. Moreover, [17] uses the content and collaborative based approaches to generate recommendation and evaluate user profiles. Wang et al. designed an approach that measures the similarity and generates a recommendation based on an inferred network tags [18].

c) Article Suggestion: Article suggestion is based on the topic and tag recommendation, which helps the user to choose the right tag. A survey of tag-based recommendation along with the evaluation was proposed in [19]. In [20], the tag was assigned automatically after analyzing the content of the article. Approaches for tagging based on correlation, rankboost, and neighbor voting graph was proposed in [21], [22], and [23] respectively.

2.2 Crowdsensing

Crowdsensing helps in reducing the overhead of data collection and processing by using mobile phones. Farkas et al. designed an application that provides information about public transport using participatory sensing [26]. The application crowdsourced data collection and feedback for visualizing the actual position of the vehicle. An end-to-end participatory noise mapping system titled as EarPhone, was proposed in [27] that uses compressive sensing for noise map recovery and outsource the environmental data collection. In [28], the authors develop an Adverse Drug Reactions (ADR) system that collects the ADR report through questions. Then apply machine learning algorithms to automate data collection procedures and efficiently track the adverse events. Hu et al. designed SmartRoad sensing system that collects data from in-vehicle smartphones GPS sensor using participatory sensing and detects traffic regulators, traffic lights, and stop signs [29]. Similarly, Wang et al. developed an application that shares the user's location along with vehicle speed in real-time [30]. The application uses participatory sensing for identifying the traffic condition and user location.

2.3 Healthcare

With the emergence of wearable and smart technology, there is a rapid growth in personalized healthcare management. A personalized wellness service recommendation system was proposed in [31], which monitors and quantifies the user activities using mobile phone sensors. In [1], the authors presented cyber-physical recommendation system that monitors user enjoyment while playing exergames. The system learns from user behavior and recommends similar games. Moreover, Dharia et al. proposed PRO-Fit framework [2], which monitors user activity and timetable using accelerometer and user's calendar respectively. Based on this information recommend a workout activity.

3 Personalization and Privacy Literacy

As from the literature, we have identified that most of the research work proposed different algorithms, frameworks, and applications, which collects the user's identifiable information along with the user's activities and behaviors. Based on this information, a personalized suggestion is generated. According to the best of our knowledge, none of the research work in the specified scope (recommender system, crowdsensing, and healthcare) has considered the effect of personalization on privacy. However, few of the studies [4, 5, 11, 16] have developed a personalized recommendation application by integrating functional interactions and user's privacy preferences. In [3], the authors have described the tradeoff between personalization and privacy, targeting online social networks only. Therefore, the purpose of our research is to create awareness among the community, because these applications not only monitor user behavior using mobile phone sensors, but also access sensitive resources such as storage, camera, microphone, and contacts. Any malicious attempt may lead to serious consequences.

In order to understand the level of personalization and privacy literacy, we have created a questionnaire using Google Forms and share the links with the students using the university mailing list. The responses were collected anonymously and allowed to use for research purpose only. Table 1 shows the demographic information of the participant. A total of 196 students has completed the survey, which includes 103 males and 93 females in the young (16 to 22), early adult (22 to 35), and middle adult (35 to 50) groups. These participants are enrolled in undergraduate, graduate, and post-graduate studies. They belong to diverse nationalities, which includes South Korea (124), China (02), Nepal (01), India

Table 1. Demographic details of Participants

		Participants
Gender	Male	103
	Female	93
Age	16 to 22	117
	22 to 35	66
	35 to 50	13
Education	Undergraduate	121
	Graduate	35
	Post Graduate	13
Nationality	South Korea	124
	China	2
	Nepal	1
	India	8
	Vietnam	9
	Ecuador	5
	Pakistan	33
	Egypt	1
	Yemen	3
	Bangladesh	10

(08), Vietnam (09), Ecuador (05), Pakistan (33), Yemen (03), and Bangladesh (10).

Each participant has answered eleven questions based on their understanding. The statistic of the survey is presented in Table 2. The survey questions were classified into personalization (5), privacy (5), and willingness (1). The results show some interesting facts, almost all the participants know about personalization and privacy. Most of the participants were satisfied with the high-quality personalization, but some do not want to share their personal data with the application. According to their reviews, they have no choice because they need to get the permission in order to use the application. Moreover, participants were also very concern about their privacy over the internet and they believe that it may be exploited by the attacker. But still, most of the participants share their post or comments without any privacy preference. According to them, it is just a post or comment, which will be useless for the attacker. However, all the participants show their interest to learn about the privacy attack that may cause by the personalization aspect. Figure 1 shows the graphical representation of the overall statistical result.

4 Exploiting Personalization Vulnerabilities

Modern service based on personalization has appealed the users due to its numerous benefits. On the other hand, it also provides a new attacking surface for

Table 2. Personalization and Privacy Questionnaire

Categorization	**Survey Questions**	**Description**	**Yes**	**No**
Personalization	**SQ1**	Do you know about personalization?	193	3
	SQ2	Do you like the personalization aspect used in different websites and mobile phone application?	189	7
	SQ3	Do you like personalized recommendation feature of different applications?	146	50
	SQ4	Do you allow the third party application to use your personal data for recommendation?	110	89
	SQ5	Do you feel comfortable, if your personal data helps in generating a high quality personalized recommendation?	170	26
Privacy	**SQ6**	Do you know about privacy?	196	0
	SQ7	Do you really concern about your privacy over the internet?	177	19
	SQ8	Do you believe that activity monitoring and tracking along with personal information may lead to certain privacy threats?	152	44
	SQ9	Do you think it would be a privacy threat, if your mobile phone exchange the information (location, driving speed) with the nearby mobile phones for a specific task?	124	72
	SQ10	Do you share your posts or comments on the social media without any privacy preference?	181	15
Willingness	**SQ11**	Do you like to be aware of privacy attacks that may cause by personalization aspect?	196	0

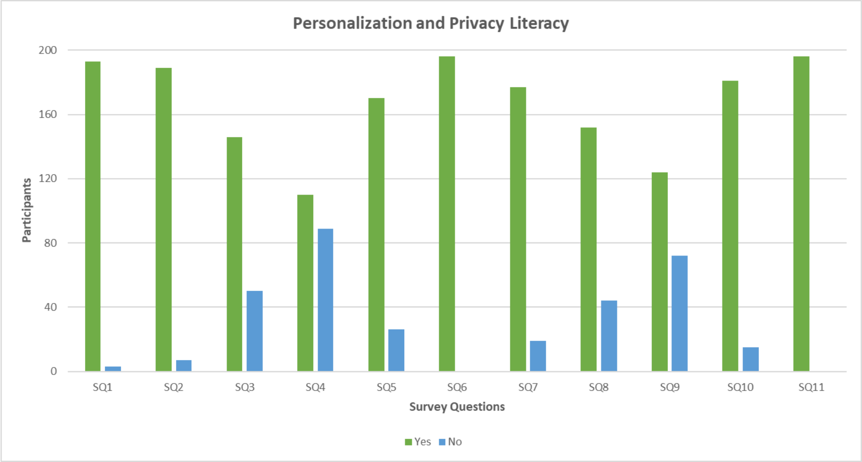

Fig. 1. Statistical Representation of Personalization and Privacy Survey Results

the attacker. In this section, we have discussed some of the attacks that exploit the personalization aspect and caused serious damage.

4.1 Pollution Attack

Pollution attack is categorized as the most effective attack for personalization services, as it allows third-party applications to modify the customize content and affect a user's choice set. In [6], Xing et al. used pollution attack and exploit the personalization aspect of YouTube, Google, and Amazon.

4.2 Data Poisoning

In data poisoning, the attacker recruited a group of malicious user to submit malicious data, which degrade the efficiency of a crowdsensing system. In [7], the authors have focused on the two types of data poisoning. *a*) Availability Attack, *b*) Target Attack

a) Availability Attack: The purpose of availability attack is to engage the malicious worker to increase the chances of error in the crowdsensing systems.

b)Target Attack: The attacker tries to skew the victim to a specific answer by poisoning the sensory data. If the victim is changed to the target answer, the attack gets to succeed.

4.3 Whaling Phishing

Whaling phishing exploits the user to reveal personal or organization information [8]. It is a special type of phishing that targets a user having sensitive information.

5 Conclusion and Future Work

In the last few years, personalization has become very popular due to its value-added services. Besides these services, it has several privacy issues, as it collects the user information based on online surfing behavior or using mobile phone sensors. The collected data have sensitive information that leads to several privacy issues. The purpose of this research is to create an awareness among the community regarding their privacy. According to the best of our knowledge, this is the first study that considered the effect of personalization on privacy, emphasizing on the emerging domains of crowdsensing and healthcare along with the recommender system. Currently, we are working on designing an algorithm that will be used by the recommender system, crowdsensing, and healthcare domains, in order to ensure personalization and privacy. Also, our design mechanism will provide prevention against the existing attacks, such as pollution, data poisoning, and whaling phishing attacks.

Acknowledgments

This research was supported by an Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (No. 2017-0-00655). This work was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-0-01629) supervised by the IITP (Institute for Information & communications Technology Promotion) and NRF-2016K1A3A7A03951968

References

1. Agu, E., Claypool, M.: Cypress: A cyber-physical recommender system to discover smartphone exergame enjoyment. In: Proceedings of the ACM workshop on engendering health with recommender systems (2016).
2. Dharia, S., Jain, V., Patel, J., Vora, J., Chawla, S., Eirinaki, M.: PRO-Fit: A personalized fitness assistant framework. In: SEKE. pp. 386–389 (2016).
3. Weinberger, M., Bouhnik, D.: Place Determinants for the Personalization-Privacy Tradeoff among Students. Issues in Informing Science and Information Technology. 15, 079–095 (2018).
4. Katragadda, B., Sharife, S.M.: Supporting Privacy Protection in Personalized Web Search. Journal of Science and Technology (JST). 2, 17–21 (2017).

5. Divekar, M.J., Patil, D.R.: Enabling Personalized Search over Encrypted Outsourced Data with Efficiency Improvement. INTERNATIONAL JOURNAL. 3, (2018).
6. Xing, X., Meng, W., Doozan, D., Snoeren, A.C., Feamster, N., Lee, W.: Take This Personally: Pollution Attacks on Personalized Services. In: USENIX Security Symposium. pp. 671–686 (2013).
7. Miao, C., Li, Q., Xiao, H., Jiang, W., Huai, M., Su, L.: Towards Data Poisoning Attacks in Crowd Sensing Systems. In: Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing. pp. 111–120. ACM (2018).
8. Bansal, G.: Got Phished! Role of Top Management Support in Creating Phishing Safe Organizations. (2018).
9. Sedhain, S., Sanner, S., Braziunas, D., Xie, L., Christensen, J.: Social collaborative filtering for cold-start recommendations. In: Proceedings of the 8th ACM Conference on Recommender systems. pp. 345–348. ACM (2014).
10. Sedhain, S., Sanner, S., Xie, L., Kidd, R., Tran, K.-N., Christen, P.: Social affinity filtering: Recommendation through fine-grained analysis of user interactions and activities. In: Proceedings of the first ACM conference on Online social networks. pp. 51–62. ACM (2013).
11. Gardner, Z., Leibovici, D., Basiri, A., Foody, G.: Trading-off location accuracy and service quality: Privacy concerns and user profiles. In: Localization and GNSS (ICL-GNSS), 2017 International Conference on. pp. 1–5. IEEE (2017).
12. Guy, I., Zwerdling, N., Ronen, I., Carmel, D., Uziel, E.: Social media recommendation based on people and tags. In: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. pp. 194–201. ACM (2010).
13. Chen, J., Geyer, W., Dugan, C., Muller, M., Guy, I.: Make new friends, but keep the old: recommending people on social networking sites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 201–210. ACM (2009).
14. Groh, G., Ehmig, C.: Recommendations in taste related domains: collaborative filtering vs. social filtering. In: Proceedings of the 2007 international ACM conference on Supporting group work. pp. 127–136. ACM (2007).
15. Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: A unified framework for providing recommendations in social tagging systems based on ternary semantic analysis. IEEE Transactions on Knowledge and Data Engineering. 22, 179–192 (2010).
16. Bhamidipati, S., Fawaz, N., Kveton, B., Zhang, A.: PriView: Personalized media consumption meets privacy against inference attacks. IEEE Software. 32, 53–59 (2015).
17. Hannon, J., McCarthy, K., Smyth, B.: Finding useful users on twitter: twittomender the followee recommender. In: European Conference on Information Retrieval. pp. 784–787. Springer (2011).
18. Wang, X., Liu, H., Fan, W.: Connecting users with similar interests via tag network inference. In: Proceedings of the 20th ACM international conference on Information and knowledge management. pp. 1019–1024. ACM (2011).
19. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in social bookmarking systems. Ai Communications. 21, 231–247 (2008).
20. Mishne, G.: Autotag: a collaborative approach to automated tag assignment for weblog posts. In: Proceedings of the 15th international conference on World Wide Web. pp. 953–954. ACM (2006).

21. Krestel, R., Fankhauser, P., Nejdl, W.: Latent dirichlet allocation for tag recommendation. In: Proceedings of the third ACM conference on Recommender systems. pp. 61–68. ACM (2009).
22. Wu, L., Yang, L., Yu, N., Hua, X.-S.: Learning to tag. In: Proceedings of the 18th international conference on World wide web. pp. 361–370. ACM (2009).
23. Freund, Y., Iyer, R., Schapire, R.E., Singer, Y.: An efficient boosting algorithm for combining preferences. Journal of machine learning research. 4, 933–969 (2003).
24. Personalization Trends. https://www.emarketer.com/Report/Personalization-Retail-Latest-Trends-Challenges/2002008
25. Jeong, Y., & Kim, Y.: Privacy concerns on social networking sites: Interplay among posting types, content, and audiences. In: Computers in Human Behavior. 69, 302-310 (2017).
26. Farkas, K., Nagy, A. Z., Tomás, T., & Szabó, R.: Participatory sensing based real-time public transport information service. In: Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on. pp. 141-144. IEEE (2014)
27. Rana, R. K., Chou, C. T., Kanhere, S. S., Bulusu, N., & Hu, W.: Ear-phone: an end-to-end participatory urban noise mapping system. In: Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks pp. 105-116. ACM (2010)
28. Chen, C., Huang, Y., Liu, Y., Liu, C., Meng, L., Sun, Y., Bian, K., Huang & Jiao, B.: Interactive crowdsourcing to spontaneous reporting of adverse drug reactions. In: Communications (ICC), 2014 IEEE International Conference on. pp. 4275-4280. IEEE (2014)
29. Hu, S., Su, L., Liu, H., Wang, H., & Abdelzaher, T. F.: Smartroad: Smartphone-based crowd sensing for traffic regulator detection and identification. ACM Transactions on Sensor Networks (TOSN), 11(4), 55 (2015).
30. Wang, C., Liu, H., Wright, K. L., Krishnamachari, B., & Annavaram, M.: A privacy mechanism for mobile-based urban traffic monitoring.In: Pervasive and Mobile Computing, 20, 1-12 (2015).
31. Afzal, M., Ali, S. I., Ali, R., Hussain, M., Ali, T., Khan, W. A., Amin, M. B., Kang, B. H., & Lee, S.: Personalization of wellness recommendations using contextual interpretation. In: Expert Systems with Applications, 96, 506-521 (2018).