Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 4, April 2017, pg.459 – 466

A REVIEW ON PRIVACY MODEL TO SECURE USER PERSONAL DETAILS IN RECOMMENDATION SYSTEM

DARSHANA D. WANKHADE, A. P. RUDEY

PG Student, Amravati University, Dr. Sau. Kamaltai Gawai Institute of Engineering & Technology, Darapur, Maharashtra, India
Assistant Professor, Amravati University, Dr. Sau. Kamaltai Gawai Institute of Engineering & Technology, Darapur, Maharashtra, India
darshanaw77@gmail.com, akshayrude@gmail.com

Abstract: Personalized recommendation has demonstrated its effectiveness in improving the problem of information overload on the Internet. However, evidences show that due to the concerns of personal privacy, users' reluctance to disclose their personal information has become a major barrier for the development of personalized recommendation. In this paper, we propose to generate a group of fake preference profiles, so as to cover up the user sensitive subjects, and thus protect user personal privacy in personalized recommendation. First, we present a client-based framework for user privacy protection, which requires not only no change to existing recommendation algorithms, but also no compromise to the recommendation accuracy. Second, based on the framework, we introduce a privacy protection model, which formulates the two requirements that ideal fake preference profiles should satisfy: (1) the similarity of feature distribution, which measures the effectiveness of fake preference profiles to hide a genuine user preference profile; and (2) the exposure degree of sensitive subjects, which measures the effectiveness of fake preference profiles to cover up the sensitive subjects. Finally, based on a subject repository of product classification, we present an implementation algorithm to well meet the privacy protection model. Both theoretical analysis and experimental evaluation demonstrate the effectiveness of our proposed approach.
Keywords: Personalized Recommendation, Personal Privacy, Sensitive Subject, Feature Distribution

I. INTRODUCTION

The rapid development of the Internet results in the explosive growth of information quantity, leading to the serious problem of information overload, and thus greatly reducing the using efficiency of information. Personalized recommendation, which can guide users to discover the information that they really need by means of the record analysis of user personal preferences, is considered to one of the most effective tools to solve the problem of information overload [1],[2].personalized recommendation has achieved great success in many application fields (typically, e-commerce). Almost all the large-scale ecommerce sites (such as Amazon and Jingdong) have introduced personalized recommendation to a variable extent. In general, a complete personalized recommendation system consists of three parts a behavior record component that collects user's personal information, a preference analysis component that analyzes user personal preferences, and a recommendation algorithm component. In a personalized recommendation system, the recommendation algorithm is the core component, which aims to find out the products hat best meet user preferences from a database of products. Presently, there exist many kinds of recommendation algorithms, typically including collaborative filtering content-based recommendation and network-based

recommendation. In general, the better the accuracy of personalized recommendation, the more users' personal information a recommendation algorithm needs to master. However, the collection and analysis of users' personal information will lead to users' concerns on personal privacy, resulting in negative impacts on the development of personalized recommendation: it not only reduces the willingness of users to use the service of personalized recommendation, but also makes users no longer willing to supply accurate personal information, thereby, reducing the accuracy of personalized recommendation. Therefore, personalized recommendation would lose the confidence and support of the users, if it cannot strengthen the protection of users' personal privacy. In fact, user privacy concerns have become one major barrier for the development and application of personalized recommendation[2].

A. Motivations

In order to protect personal privacy in personalized recommendation, many approaches have been proposed, specifically including: data obfuscation, data transformation, anonymization etc. (1) The basic idea of **data obfuscation** techniques is to use fake or general data to obfuscate the data related to the sensitive preferences contained in users' preference profiles [6] This kind of techniques might lead to poor recommendation accuracy due to its change to user preference profiles. (2) In **data transformation** techniques, users' personal data need to be transformed (e.g., using noise addition or data perturbation)[9] before being used for personalized recommendation. Generally, this kind of techniques can only be applied to collaborative filtering algorithms. Moreover, it has been demonstrated that effective data transformation would not lead to a negative impact on the accuracy of collaborative filtering recommendation. However, since the recommendation results are fully visible to the untrusted server-side, it is possible for an attacker on the server-side to guess the genuine user preferences conversely by analyzing the recommendation results, thus, leading to the disclosure of personal privacy. (3) **Anonymization** has been widely applied to personal privacy protection which allows users to use a system without the need to expose their identity information. However, as pointed out in it is very necessary to confirm the true identity for each user in a recommendation system. Therefore, this kind of techniques cannot satisfy the requirement of the practical application of personalized recommendation. Based on the above, we conclude that to supply an effective personalized recommendation service, it is required for a privacy protection approach to satisfy the following three requirements. (1) Ensuring the **security** of user sensitive preferences (i.e., the preference information that users are not willing to expose). Specifically, it should be difficult for an attacker not only to identify the user sensitive preferences from users' personal behavior (or data), but also to guess the user sensitive preferences conversely through analyzing the results returned from the recommendation algorithm. The former can be achieved by both data obfuscation and data transformation. However, the latter cannot be achieved by data transformation since it ensures the accuracy of recommendation. (2) Ensuring the **accuracy** of the user final recommendation results, i.e., the recommendation results that users receive finally should be as consistent as possible (or the same), before and after the privacy protection approach is introduced. (3) Ensuring the **efficiency** of personalized recommendation, i.e., the introduction of privacy protection should not lead to a serious effect on the execution efficiency of a personalized recommendation service.

B. Contributions

We aim to propose an effective approach to protect user's personal privacy in personalized recommendation. The approach should address all the problems mentioned above, i.e., under the precondition of not changing existing recommendation algorithms, it can not only effectively prevent the untrusted server-side from identifying the user sensitive preferences from personal data or recommendation results, but also ensure the accuracy of recommendation results and the efficiency of a personalized recommendation service. The basic idea of the approach is to construct a group of fake preference profiles, so as to cover up the user sensitive subjects, and thus to protect user personal privacy. Specifically, the contributions of this paper are threefold. First, we present a client-based system framework to protect user sensitive preferences in personalized recommendation. Under the system framework, we move the behavior record component to a trusted client, making that user preference profiles would be generated in the trusted client. Then, the client constructs a group of fake preference profiles, and submits them together with the genuine user preference profile to the server-side for personalized recommendation. Thus, the recommendation results from the server-side would be no longer accurate (since including those corresponding to the fake profiles), which makes it difficult for an attacker to identify the user's sensitive preferences from the recommendation results. Finally, the client discards all the recommendation results that correspond to the fake preference profiles, so only the recommendation result that corresponds to the genuine preference profile is returned to the user, consequently, ensuring the accuracy of personalized recommendation. Second, based on the system framework, the paper introduces a privacy model for user sensitive preference protection. The model formulates the requirements that the fake preference profiles should satisfy so as to protect the sensitive preferences effectively, i.e., fake profiles should have similar features with the genuine profile, and irrelevant subjects with the sensitive preferences. The feature similarity makes it difficult for an attacker to identify the genuine user preference profile, even if the attacker captures all the preference profiles. The subject irrelevance results in that the exposure degree of the sensitive preferences on the server-side can be effectively reduced by the fake profiles, thereby, ensuring the security of users' sensitive preferences. Finally, according to the system framework and the privacy model mentioned above,

based on a subject repository of product classification, we present an implementation algorithm that runs on a trusted client. The algorithm can well meet the requirements of user privacy protection in personalized recommendation, i.e., it can construct a group of fake preference profiles that well meet the privacy model. In addition, we have demonstrated the effectiveness of the privacy model and its implementation algorithm through theoretical analysis and experimental evaluation.

II. LITERTURE REVIEW AND RELETED WORK

Depending on the recommendation algorithms, recommendation systems can be divided into three main categories: (1) collaborative filtering which is the process of filtering products based on the similarity computation of users' previous preference products; (2) content-based recommendation which recommends products for a user based on the similarity between the user preferences and the product descriptions; and (3) social network based recommendation which is an extension of collaborative filtering, and measures the similarity of users using a social network analysis technique. In general, a recommendation algorithm has to run on an untrusted server-side, and the better the recommendation accuracy, the more users' personal information the algorithm needs to master, consequently, leading to users' serious concerns on personal privacy In order to protect user privacy in personalized recommendation, many approaches have been proposed. In this section, we briefly review and analyze these approaches, specifically, including: data obfuscation, data transformation, anonymization etc.

A. Data Obfuscation

The basic idea of data obfuscation techniques is to leverage fake data or general data to obfuscate the data related to the sensitive preferences contained in user preference profiles. In order to protect the genuine intention hidden in a user query, into the user query. Then, similar approaches are also proposed but they allow a user to define his own privacy requirements, i.e., to define the subjects that the user wants to protect, and the degree of protection. Aiming at personalized advertisement recommendation, a client-based approach to user privacy protection, which is based on the comprehensive consideration of user privacy (i.e., the privacy level that a user is willing to share with the server-side) and network traffic (i.e., the number of ads returned to a mobile phone) to select relevant ads for a user. Aiming at personalized web search, It builds a hierarchical structure of user preferences on the client, where nodes of high level are used to store general preference subjects, while other nodes of low level are used to store special subjects. Then, some general subjects are selected to replace sensitive special subjects, so as to protect the user sensitive preferences. which also propose to cover up the user interested preferences using more general preferences. However, this kind of techniques certainly will reduce the recommendation accuracy due to its change to user preference profiles, namely, whose privacy protection is based on a compromise on recommendation performance.

B. Data Transformation

In data transformation techniques, users' personal data need to be transformed (e.g., by noise addition or data perturbation)before being used for personalized recommendation. Generally, this kind of techniques can only be applied to collaborative filtering algorithms. Random perturbation technique (RPT) is a frequently-used approach for data transformation. Its basic idea is to attach a random data (r) to the user sensitive data (a) so that what an attacker can see is ($a + r$), i.e., submit the user sensitive data together with the additional random data to the server for personalized recommendation, so that the server cannot see the true user data. When the user data quantity is large enough, by using the overall user data for collaborative filtering recommendation, we can still obtain a relatively accurate recommendation result. Thus, RPT can ensure not only the security of user privacy, but also the recommendation accuracy. A similar method is proposed in to protect the personal privacy of data mining. The paper [21] designs a collaborative filtering recommendation system based on the discrete wavelet transform (DWT) technique and random perturbation technique. The paper [19] proposes to write several well-designed "predictive scores" into a user-product scoring matrix (that is the input of a collaborative filtering algorithm), so as to perturb the true user scoring information and thus protect personal privacy. The paper [30] has evaluated the effect of data transformation on the accuracy of collaborative filtering recommendation. The results show that effective data transformation would not lead to a negative impact on the accuracy of collaborative filtering recommendation. It can be seen that this kind of techniques can ensure not only the accuracy of recommendation results to a certain extent, but also the security of a sensitive preference in its user preference profile effectively. However, the accuracy of a recommendation result leads to that many products relevant to the user sensitive preferences are generally contained in the recommendation result. Since the recommendation result is fully visible to the untrusted server-side, it is possible for an attacker on the server-side to guess the genuine user preferences conversely through analyzing the recommendation result, consequently, leading to the disclosure of personal privacy.

C. Anonymization

Anonymization is a kind of widely used approaches in privacy protection. It allows users to use a system without the need to expose their identity information. Anonymization, due to the non complexity of its processing, can be easily applied to a personalized recommendation system, and has been widely used in many systems to protect user personalprivacy, such as However, there have been many questions about the practicality of using anonymization for privacy protection in personalized recommendation. present the shortages of anonymization to user privacy protection, and demonstrate the results by using experiment evaluations. Anonymization increases the possibility that a user submits useless random data, thereby, decreasing the quality of user personal data. Moreover, anonymization also makes the system easier to be attacked by competitors. For example, a company can submit a large number of fake data in a recommendation system to promote its own products to obtain more opportunities of recommendation. Thus, it is necessary to confirm the true identity for each user in a recommendation system. At present, most of personalized recommendation systems require users to provide the basic information that can identify their personal identities. Therefore, this kind of techniques cannot satisfy the requirement of the practical application of personalized recommendation

III. ANALYSIS OF PROBLEM

We study an approach for protecting user sensitive subjects in a personalized recommendation system. According to the motivations presented in Section 1.1, the approach has to meet the following four requirements. (1) It does not change the existing structure of a recommendation algorithm. (2) It does not compromise the accuracy of the final recommendation. (3) It ensures the security of user preferences, making it difficult for an attacker not only to identify the sensitive subjects from a user preference profile, but also to guess the sensitive subjects conversely from the recommendation result. (4) It does not lead to a serious effect on the execution efficiency of a personalized recommendation service. In this section, we present the system model used in our approach, and then discuss the attack model based on the system model.

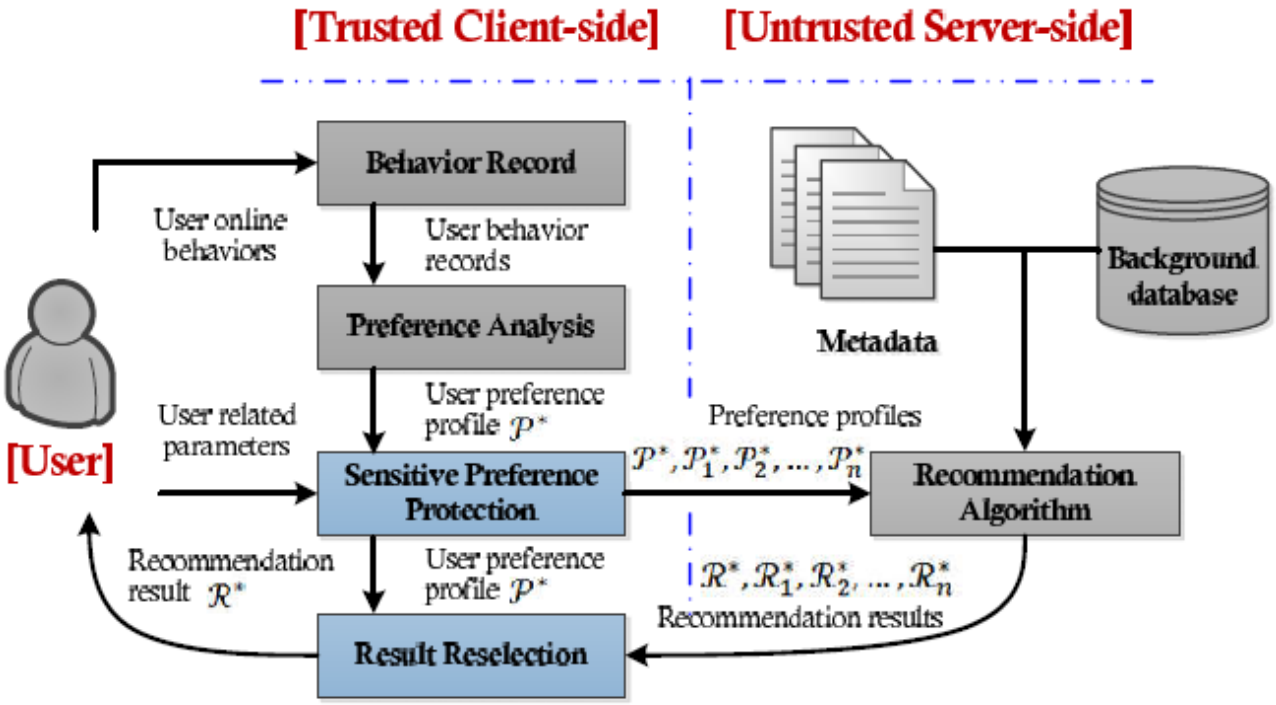


Fig.1. The system framework for the protection of user sensitive preferences in a personalized recommendation service, where the blue components sensitive *preference protection" and "result reselection" are introduced newly.*

A. System Model

Here, user sensitive preferences are referred as the personal preferences that users are unwilling to be seen or analyzed by attackers. Fig. 1 shows the system framework used by this paper for the protection of user sensitive preferences in a personalized recommendation service, which consists of an untrusted server-side and many trusted client-sides. The basic process flow of the system framework is presented as follows.

• Under the client-based architecture, the user behavior record component and the preference analysis component are moved from the server to a client. Thus, the client (instead of the server) collects and analyzes user behaviors to generate a user preference profile $P*$.

• In the client, the newly-introduced component of sensitive preference protection constructs a group of fake preference profiles $P*;P*;:::;P*n$ based on the user preference profile $P*$, after taking into consideration the requirements of security, accuracy and efficiency. Then, the fake preference profiles are submitted together with the genuine user preference profile to the server-side, as the input of the personalized recommendation algorithm.

• In the client, the newly-introduced result reselection component selects the recommendation result $R*$, which corresponds to the user preference profile $P*$, from all the recommendation results $R*;R*;R*;:::;R*n$ that are returned by the recommendation algorithm on the server-side. Then, the component returns $R*$ to the user, while discarding the other recommendation results $R*;R*;:::;R*n$. Based on the system framework in Fig. 1, we conclude as follows. On the one hand, the results outputted by the recommendation algorithm component in the server-side, are no longer equal to the true user recommendation result (i.e., the result before the introduction of privacy protection). They contain the recommendation results corresponding to the fake preference profiles. Thus, it is difficult to immediately identify the user sensitive preferences from the recommendation results. On the other hand, the results outputted by the recommendation algorithm are certainly a superset of the true recommendation result, thereby ensuring that the user can obtain an accurate recommendation. In addition, the system framework requires no change to the existing personalized recommendation algorithm, so it is transparent for both the user on the client and the recommendation algorithm component running on the server-side. However, from Fig. 1, it can also be seen that the fake preference profiles generated by the component of sensitive preference protection play an important role in the framework, i.e., their quality is the key to user privacy protection. Generally, the fake preference profiles generated randomly are easy to be ruled out, thus failed to cover up the sensitive preferences contained in a user preference profile. This is because the features of user preferences are generally regularly distributed (e.g., a user is interested in one or several fixed subjects for a period of time), while randomly generated preference profiles are not (which may be evenly related to a large number of subjects). Thus, an attacker can easily detect fake preference profiles according to their different feature distribution. In addition, the fake preference profiles should be not related to the user sensitive preferences. For example, suppose that a sensitive preference related to a user preference profile is the subject "sporting goods". Then it is not appropriate to generate a group of fake profiles that also contain the sensitive subject "sporting goods" or other highly relevant subjects, because at this time, an attacker can immediately draw a conclusion that the user is interested in "sporting goods", without ruling out the fake profiles. To this end, fake preference profiles generated by the sensitive preference protection component should meet the following two requirements: (1) ensuring the security of user sensitive preferences on the untrusted server-side, i.e., reducing the exposure degree of user sensitive preferences on the server side, and hence the probability of an attacker to detect them; and (2) exhibiting highly-similar feature distribution with the user preference profile, so as to make it difficult for an attacker to rule out the fake profiles, thus, hiding the user profile effectively.

B. Attack Model

In the system framework, the server-side is not trusted, which is considered as the biggest potential attacker. Assume that the attacker has taken control of the server (i.e., the attacker may be a hacker who breaks the server, or an administrator who works on the server). Thus, the proposed approach to user privacy protection needs to prevent the server from identifying the sensitive preferences related to a user preference profile. From the system framework shown in Fig. 1, we can see that the attacker can obtain not only all the preference profiles submitted by the client, but also all the recommendation results generated by the personalized recommendation algorithm. Thus, we need to prevent the attacker from identifying the user sensitive preferences not only from the preference profiles, but also from the recommendation results. In addition, because of taking control of the server, the attacker has a powerful capability, which masters the database of all the products and the repository of product classification, and takes charge of executing the personalized recommendation algorithm. Unfortunately, the attacker might also know the existence of the sensitive preference protection algorithm deployed on the client, and obtain a copy of the algorithm. Hence, the attacker can input each of the mastered preference profiles to the privacy protection algorithm, and then observe the output results to guess the user preference profile.

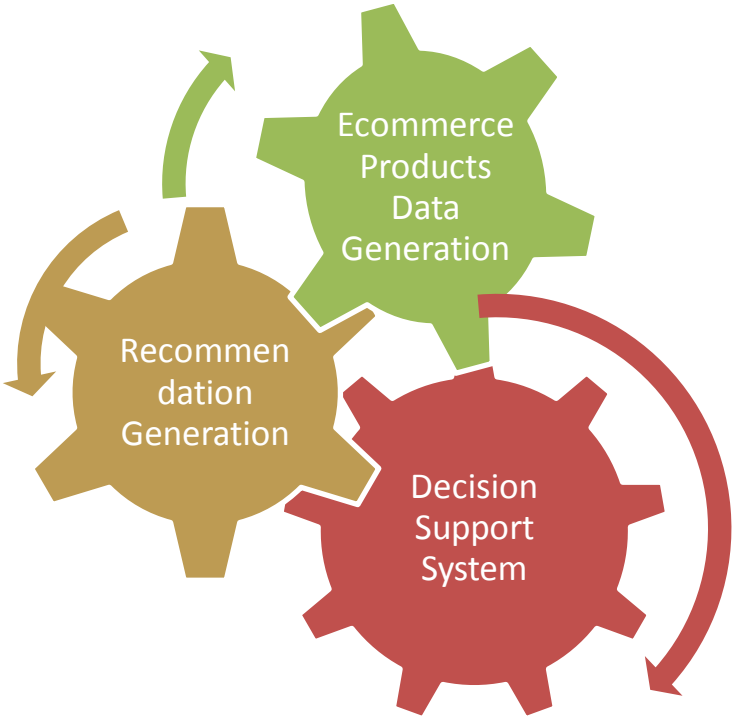
Symbols	Meanings
$\mathcal{P}$	A set of all the products
$\mathcal{P}^*$	A set of user preference products, i.e., $\mathcal{P}^* \subseteq \mathcal{P}$
$\mathcal{G}$	A set of all the subjects
$\mathcal{G}^*$	A set of user preference subjects, i.e., $\mathcal{G}^* \subseteq \mathcal{G}$
$\mathcal{G}_k^*$	A set of user preference subjects with the level k, i.e., $\mathcal{G}_k^* \subseteq \mathcal{G}^*$
$\mathcal{G}^\dagger$	A set of user sensitive preference subjects, i.e., $\mathcal{G}^\dagger \subseteq \mathcal{G}^*$
k^m	The maximum of levels for all the subjects, i.e., $k^m = \max_{g \in \mathcal{G}}\{level(g)\}$
P	The product feature distribution vector, which corresponds to $\mathcal{P}^*$
G^k	The subject feature distribution vector, which corresponds to $\mathcal{G}_k^*$

Table 1. Symbols and their meanings

IV. PROPOSED WORK AND OBJECTIVES

We proposed an approach for protecting personal privacy for users when using a personalized recommendation service, whose basic idea is to construct a group of fake preference profiles to cover up the sensitive subjects contained in a user preference profile, and in turn protect user personal privacy. We used a client-based system framework that requires not only no change to the existing recommendation algorithms, but also no compromise to the accuracy of recommendation results. Finally, both theoretical analysis and experimental evaluation have demonstrated the effectiveness of our approach: (1) it can generate a group of good-quality fake preference profiles, which not only have high feature distribution similarities with the genuine user preference profile (so as to hide the genuine profile), but also can be used to effectively reduce the risk of exposing the user sensitive subjects; and (2) it does not cause serious performance overheads on either running time or running memory. Therefore, we conclude that our approach can be used to effectively protect users' personal privacy in personalized recommendation.

- To develop a security model which maintains user's personal details secure against various attacks
- To develop a product recommendation system which recommends user's preferences wise ecommerce products
- To implement Data Obfuscation, Data Transformation and Anonymization techniques to enhance data security



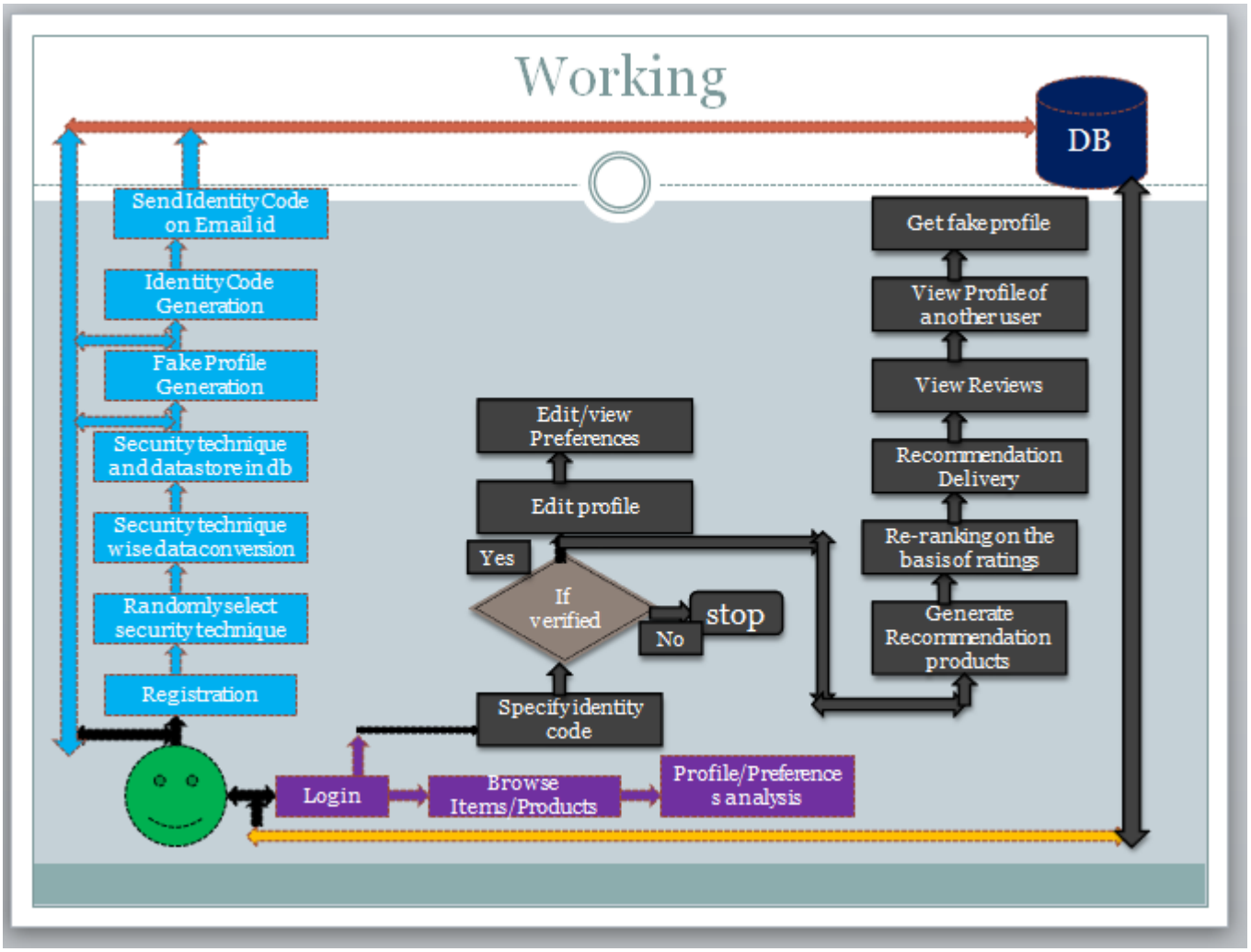


Fig.3. Privacy model to secure user personal details in recommendation system working

V. CONCLUSION

We proposed an approach for protecting personal privacy for users when using a personalized recommendation service, whose basic idea is to construct a group of fake preference profiles to cover up the sensitive subjects contained in a user preference profile, and in turn protect user personal privacy. We used a client-based system framework that requires not only no change to the existing recommendation algorithms, but also no compromise to the accuracy of recommendation results. Finally, both theoretical analysis and experimental evaluation have demonstrated the effectiveness of our approach: (1) it can generate a group of good-quality fake preference profiles, which not only have high feature distribution similarities with the genuine user preference profile (so as to hide the genuine profile), but also can be used to effectively reduce the risk of exposing the user sensitive subjects; and (2) it does not cause serious performance overheads on either running time or running memory. Therefore, we conclude that our approach can be used to effectively protect users' personal privacy in personalized recommendation.

ACKNOWLEDGEMENT

We would like to acknowledge the Faculties of Computer Science and Engineering Department. Dr. Sau. Kamaltai Gawai institute of engineering & Technology, Darapur, Amravati for their support. I Miss Darshana D. Wankhade specially want to thank my guide Prof. A.P. Rude for their guidance and constant encouragement towards the project work. . I would like to thank to respective teachers for their constant and valuable support and motivation. Last but not least, I would like to thank all who directly or indirectly helped me in processing the paper.

REFERENCES

[1] Zongada wu, Guiling Li, Qi Liu Guandong Xu, and Enhong Chen et al. "Covering the Sensitive Subjects to Protect Personal Privacy in Personalized Recommendation", IEEE Transactions on Services Computing , 2016,DOI 10.1109

[2] Adem Ozturk, Huseyin Polat. "From existing trends to future trends in privacy-preserving collaborative filtering". Data Miningand Knowledge Discovery, 2015, 5 (6): 276–291

[3] A. B. Barragans-Martinez, E. Costa-Montenegro, J.C. Burguillo et al. "A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition". Information Sciences, 2010, 180 (22): 4290–4311

[4] Silvia Puglisi , Javier Parra-Arnau , Jordi Forn et al. "On content based recommendation and user privacy in social-tagging systems".Computer Standards & Interfaces, 2015, 41: 17–27

[5] Carrer-Neto, Marła Luisa Hernndez-Alcaraz, Rafael Valencia-Garcła et al. "Social knowledge-based recommender system. Application to the movies domain". Expert Systems with Applications,2012, 39 (12):10990–11000

[6] HweeHwa Pang, Xiaokui Xiao, Jialie Shen. "Obfuscating the topical intention in enterprise text search". Proc. of IEEE International Conference on Data Engineering (ICDE), 2012, pp. 1168–1179

[7] Zhifeng Luo, Shuhong Chen, Yutian Li. "A distributed anonymization scheme for privacy-preserving recommendation systems", Proc. of IEEE Conference on Software Engineering and

Service Science (ICSESS), 2013, pp. 491–494

[8] Yilin Shen, Hongxia Jin. "Privacy-preserving personalized recommendation: An instance-based approach via differential privacy".Proc. of IEEE Conference on Data Mining (ICDM), 2014, pp. 540–549

[9] Huseyin Polat, Wenliang Du. "Privacy-preserving collaborative filtering using randomized perturbation techniques". Proc. of IEEE Conference on Data Mining (ICDM), 2003, pp. 625–628

[10] Zafer Duzen, Mehmet S. Aktas." An Approach To Hybrid Personalized Recommender Systems", IEEE Journel, 2016, 978-1-4673-9910-4

