

Privacy-preserving and yet Robust Collaborative Filtering Recommender as a Service

No Author Given

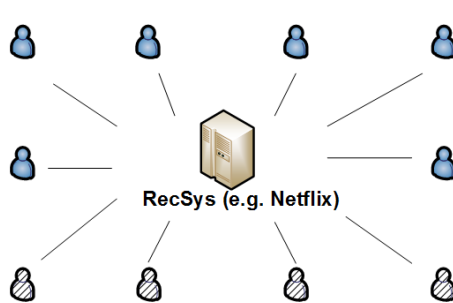
No Institute Given

Abstract. Collaborative filtering recommenders provide effective personalization services at the cost of sacrificing the privacy of their end users. Due to the increasing concerns from the society and stricter privacy regulations, it is an urgent research challenge to design privacy-preserving and yet robust recommenders which offer recommendation services to privacy-aware users. Our analysis shows that existing solutions fall short in several aspects, including lacking attention to the precise output to end users and ignoring the correlated robustness issues. In this paper, we provide a general system structure for latent factor based collaborative filtering recommenders by formulating them into *model training* and *prediction computing* stages, and also describe a new security model. Aiming at pragmatic solutions, we first show how to construct privacy-preserving and yet robust *model training* stage based on existing solutions. Then, we propose two cryptographic protocols to realize a privacy-preserving *prediction computing* stage, depending on whether or not an extra proxy is involved. Different from standard Top-k recommendations, we alternatively let the end user retrieve the unrated items whose predictions are above a threshold, as a result of our *privacy by design* strategy. Experimental results show that our new protocols are quite efficient.

1 Introduction

Today, personalization is widely adopted by a large number of industries, from entertainment to precision medicine. The main enabling technology is recommender systems, which employ all sorts of techniques to predict the preferences of human subjects (e.g. the likes and dislikes towards a movie). A typical system architecture is shown in Figure 1.

So far, a lot of generic recommender algorithms have been proposed, as surveyed in [24]. Recently, deep learning has become a very powerful tool and has been used to numerous applications, including recommender [32]. Nevertheless, the collaborative filtering recommender systems are most popular and well-known due to their explainable nature (e.g. you like x so you may also like y). Given a user set $\mathcal{U} = \{1, 2, \dots, N\}$ and their rating vectors $\mathbf{R}_i$ for $i \in \mathcal{U}$, let $\mathcal{R}$ denote the set of (i, j) such that $r_{i,j} \neq 0$. One of the most popular collaborative filtering algorithms is based on low-dimensional factor models, which derive two feature matrices $\mathbf{U}$ and $\mathbf{V}$ from the rating matrix. The feature vector $\mathbf{u}_i$ denotes



Real-world recommenders often build a preference model based on data from a number of sources, such as user’s explicit feedback (e.g. rating vectors) and implicit information (e.g. how long a user has stayed on the page of an item). Like in most literature work and for simplicity reasons, we only consider explicit feedback in this paper. The discussions and proposed solutions might also be applied to other types of data.

Fig. 1. Standard Recommender Structure

user i ’s interest and the feature vector $\mathbf{v}_j$ denotes item j ’s characteristics. Every feature vector has the dimension k , which is often a much smaller integer than M and N . In implementations, $\mathbf{U} = \{\mathbf{u}_i\}_{1 \leq i \leq N}$ and $\mathbf{V} = \{\mathbf{v}_j\}_{1 \leq j \leq M}$ are often computed by minimizing the following function:

$$\min_{\mathbf{U}, \mathbf{V}} \frac{1}{|\mathcal{R}|} \sum_{(x,j) \in \mathcal{R}} (r_{x,j} - \langle \mathbf{u}_x, \mathbf{v}_j \rangle)^2 + \lambda \sum_{x \in \mathcal{U}} \|\mathbf{u}_x\|_2^2 + \mu \sum_{j \in \mathcal{I}} \|\mathbf{v}_j\|_2^2 \quad (1)$$

for some positive parameters λ, μ , typically through the stochastic gradient descent (SGD) method or its variants. Note that one advantage of the latent factor based collaborative filtering is its better resistance to robustness attacks than the neighbourhood-based ones [20].

1.1 Privacy and Robustness Issues

Besides the likes and dislikes, users’ preferences might lead to inferences towards other sensitive information about the individuals, e.g. the religion, political orientation, and financial status. When a user is involved in a recommender system with a pseudonym, there is the risk of re-identification. For instance, Weinsberg et al. [30] demonstrated that what has been rated by a user can potentially help an attacker identify this user. Privacy issues have been recognized for a long time and a lot of solutions have been proposed today, as surveyed in [5,14]. Robustness is about controlling the effect of manipulated inputs, and is a fundamental issue for recommender systems. Its importance can be easily seen from the numerous scandals, including fake book recommendations ¹, fake phone recommendations ² and malicious medical recommendations ³. In their seminal work, Lam and Riedl [17] investigated the concept of shilling attacks, where a malicious company lies to the recommender system (or, inject fake profiles) to have its own products recommended more often than those from its competitors. Following

¹ <https://tinyurl.com/y9nyo8y9>

² <https://tinyurl.com/ycc8lujh>

³ <https://tinyurl.com/ybuevrvq>

this, a number of works have been dedicated to the investigation of different robustness attacks and corresponding countermeasures. Interestingly, Sandvig, Mobasher, and Burke [20] empirically showed that model-based algorithms are more robust than memory-based algorithms; Cheng and Hurley [10] proposed informed model-based attacks against trust-aware solutions, and demonstrated it against the privacy-preserving solution by Canny [8].

Clearly, robustness attacks pose a threat to the business perspective of the RecSys and subsequently impact the quality of service for the users. Privacy is increasingly becoming a concern for the privacy-aware users, and it is also a concern for the RecSys when it wants to deploy a *machine learning as a service* business model [28]. Unfortunately, privacy and robustness have a *complementary yet conflicting* relationship. On the complementary side, it is clear that privacy disclosure can lead to more successful robustness attacks as the attacker can adapt its attack strategy accordingly, and a robust system reduces the attack surface for the privacy attackers who injects fake profiles to infer the honest users' information based on the received outputs. On the conflicting side, a privacy-preserving recommender makes it harder to combat robustness attacks because the robustness attack detection algorithms will not work well when all users' inputs are kept private. We elaborate on this aspect in Section 3.

1.2 Our Contribution

In this paper, we aim at a comprehensive investigation of the privacy and robustness issues for recommender systems, by considering both the *model training* and the *prediction computing* stages. To this end, we first provide a general system architecture and present a high-level security model accordingly. We then review the existing privacy-preserving latent factor based recommender solutions and identify their potential issues. Particularly, we notice that most cryptographic solutions have mainly aimed at the privacy protection for the *model training* stage without paying much attention to the *prediction computing* stage. This consequently results in serious privacy issues in practice. We also highlight that existing privacy-preserving solutions make it harder to detect and prevent robustness attacks.

Towards privacy-preserving solutions that respect robustness attack detection, we separately address the issues in the *model training* and *prediction computing* stages. For the former, we show that existing solutions can be adapted, particularly it is straightforward for the expert-based ones such as that from [29]. As to the latter, we propose two new cryptographic protocols, one of which involves an extra proxy. Our experimental results show that both protocols are very efficient with respect to practical datasets. The employed *privacy by design* approach, namely returning unrated items whose approximated predictions are above a threshold, might have profound privacy implications, nevertheless we leave a detailed investigation as future work.

1.3 Organisation

The rest of the paper is organised as follows. In Section 2, we introduce a generic recommender system architecture that consists of two stages: model training and prediction computing. Accordingly, we present a high-level security model. In Section 3, we analyse some representative privacy-preserving recommender solutions and identify their deficiencies in our security model. In Section 4, we present a solution framework to demonstrate how to construct secure recommender solutions in our security model. In Section 5, we propose a new privacy-preserving protocol for prediction computing, which does not involve a third-party proxy. In Section 6, we propose a new privacy-preserving protocol for prediction computing, which is more efficient but relies on a proxy. In Section 7, we conclude the paper.

2 System Architecture and Security Model

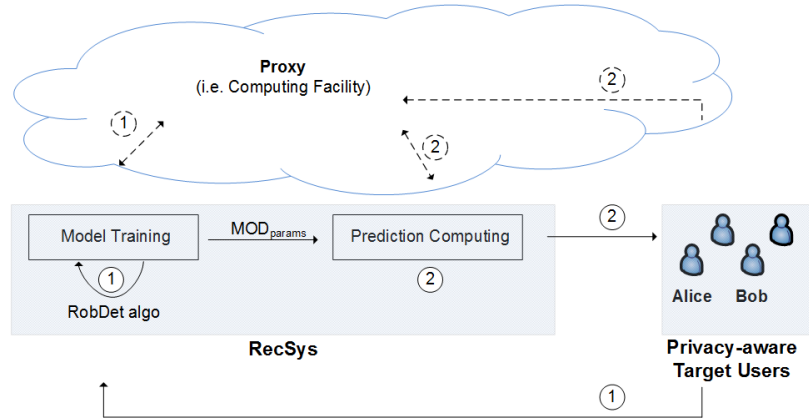


Fig. 2. Recommender as a Service Architecture

We assume the RecSys builds recommender models and offers recommendation as a service to the users. If some users do not care about their privacy, then they can offer their rating vectors directly to the RecSys to receive recommendations. In addition, the Recsys may collect as much non-private data as possible in order to build an accurate recommender model. We assume there are privacy-aware users who are not willing to disclose their rating vectors while still wishing to receive recommendations. Our main objective is to design solutions to guarantee that, from the view point of a privacy-aware user Alice,

- She receives high-quality recommendations, by avoiding the robustness attacks mentioned in Section 1.1.

- She minimizes the information disclosure about her rating vector, under the prerequisite that she receives high-quality recommendations.

For our recommender as a service, we assume a system architecture shown in Figure 2. We note that existing collaborative filtering recommender systems typically have the *model training* and *prediction computing* stages even though they might not mention them explicitly. In addition, it is also quite often that a proxy (i.e. cloud computing facility) is employed to carry out the massive computations (e.g. Netflix heavily uses Amazon cloud services). It is worth emphasizing that many privacy-preserving solutions (particularly cryptographic solutions) also introduce such a third party, e.g. the crypto service provider in the term of [16] and [21]. For the different usage scenarios, the trust assumptions on the proxy can vary a lot, and we elaborate on it later. Next, we briefly introduce what will happen in the two stages.

1. In the *model training* stage, labeled in Figure 2, the RecSys trains a model, e.g. similarities between items (or users) in neighbourhood-based recommenders and feature matrices for users and items in latent model based ones, based on data from one or more sources. To clean the data and detect robustness attacks, before the training, we suppose that the RecSys will run an algorithm RobDet over the training dataset. To simplify our discussion, we assume the the output of RobDet is a binary bit for every input profile (i.e. rating vector). If it is 0, then the profile is deemed as malicious so that will not be used in the training.
2. After training, we refer to the output of the model training stage as a set of parameters MOD_{params} . Note that the parameters might be in an encrypted form when privacy protection has been applied. In the *prediction computing* stage, the RecSys uses the model parameters MOD_{params} and possibly Alice’s rating vector to infer Alice’s preferences.

2.1 The Proposed Security Model (high level)

We make the following general assumptions related to security. First of all, we assume the communication channel is secured with respect to confidentiality and integrity in the sense: (1) an honest user can be assured that his input will reach the RecSys or another intended party without being eavesdropped on and manipulated; (2) the RecSys can be assured that the honest user, who initiates the communication, will receive the message without being eavesdropped on and manipulated. It is worth stressing that there is no guarantee whether RecSys knows the true identity of the user it is communicating with. Secondly, we assume that the RecSys is a rational player and offers recommendation as a service and a user offers monetary rewards for receiving recommendations. Without this assumption, there will not be any guarantee for achieving privacy and robustness because the RecSys will deviate from the protocol for any possible benefits.

Regarding robustness, we require that the RecSys is able to (efficiently) run any chosen RobDet algorithm over the training dataset to identify the malicious

profiles, i.e. rating vectors, as we have described in the beginning of this section. The output of RobDet should be the same regardless what privacy protection mechanisms have been deployed.

Regarding privacy, we consider the following specific requirements. Note that, similar to the semantic security of encryption schemes, indistinguishability-based games can be defined to formally capture all requirements. We skip the details here, partially due to the fact that the cryptographic primitives we use (e.g. homomorphic encryption) guarantee indistinguishability straightforwardly.

- *Alice’s privacy against RecSys.* If the RecSys does not collude with the proxy, then it learns nothing about Alice’s input and output except for information implied in the output of RobDet (i.e. whether or not Alice’s profile is suspicious if it has been used in the model training stage).
- *Alice’s privacy against Proxy.* If the proxy does not collude with the RecSys, then it learns nothing about Alice’s input and output.
- *Alice’s privacy against other users.* Other users do not learn more information about Alice’s rating vector than that implied in the legitimate outputs they receive.
- *RecSys’s privacy against Alice and other users.* Alice and other users do not learn more information than that implied in the legitimate outputs they receive.

As a remark, in many existing solutions reviewed in Section 3.2, the legitimate outputs can contain too much private information. This has motivated our privacy-by-design approach in Section 4.1. As an informal requirement, when both the RecSys and the Proxy are compromised simultaneously, the information leakage about the privacy-aware users’ data should also be minimized. To this end, we note that most existing solutions except for the expert-based ones will leak everything.

3 Literature Work and Standing Challenges

Regardless efficiency, designing a secure recommender system is a very challenging task. For example, applying statistical disclosure mechanisms does not guarantee security, as Zhang et al. [31] showed how to recover perturbed ratings in the solutions by Polat and Du [23]. Employing advanced cryptographic primitives is also not a panacea, as Tang and Wang [27] pointed out a vulnerability in the homomorphic encryption based solution by Jeckmans et al. [15]. Next, we analyse some representative solutions from the literature and identify the standing challenges.

3.1 Preliminary on Building Blocks

We use the notation $x \stackrel{\$}{\leftarrow} Y$ to denote that x is chosen from the set Y uniformly at random. A public key encryption scheme consists of three algorithms

(Keygen, Enc, Dec): Keygen(λ, L) generates a key pair (PK, SK) ; Enc(m, PK) outputs a ciphertext c ; Dec(c, SK) outputs a plaintext m . Some schemes, e.g. Paillier [22], are additively homomorphic, which means there is an operator $\oplus$ such that $\text{Enc}(m_1, PK) \oplus \text{Enc}(m_2, PK) = \text{Enc}(m_1 + m_2, PK)$. While some recent somewhat homomorphic encryption (SWHE) schemes are both additively and multiplicatively homomorphic to a certain number of operations, which means there are operators $\oplus$ and $\otimes$ such that $\text{Enc}(m_1, PK) \oplus \text{Enc}(m_2, PK) = \text{Enc}(m_1 + m_2, PK)$ and $\text{Enc}(m_1, PK) \otimes \text{Enc}(m_2, PK) = \text{Enc}(m_1 m_2, PK)$. In practice, one of the most widely-used SWHE library is Simple Encrypted Arithmetic Library (SEAL) from Microsoft [11], which is an optimized implementation of the YASHE scheme [6]. Note that homomorphic subtraction $\ominus$ can be directly defined based on $\oplus$ and with similar computational cost.

3.2 Examining some Cryptographic Solutions

Cryptographic solutions aim at minimizing the information leakage in the computation process, by treating the recommender as a large-scale multi-party computation protocol. When designing privacy-preserving solutions, it has become a common practice to introduce one or multiple third parties, not all of which are supposed to collude with each other, in order to eliminate a single trusted third party and improve efficiency. Nikolaenko et al. [21] and Kim et al. [16] introduced a CSP (i.e. crypto service provider) and employed garbled circuits and homomorphic encryption respectively to perform privacy-preserving matrix factorization. These solutions put the emphasis on the matrix factorization step (i.e. model training stage) while failing to pay more attention to the prediction computing stage. In [21], it is proposed that every user i is given his own feature vector $\mathbf{u}_i$ so that it can interact with the RecSys and CSP to retrieve predictions on all items (i.e. $\mathbf{u}_i(\mathbf{v}_j)^T$ ($1 \leq j \leq M$)). In reality, the users do not need to know his feature vector and the predictions to all items, they only need to know the items they might like. In more detail, there are several concerns.

- Given the fact that $M \ll N$ (i.e. the number of items are far less than the user population), a small number of colluded users can recover the item feature matrix $\mathbf{V}$, based on which they can try to infer information about the rest of the population. This leads to unnecessary information leakages against the honest users.
- The malicious users might make illegal use of the recovered $\mathbf{V}$, through providing recommendation services using technologies, such as incremental matrix factorization. Besides the potential privacy concern, this may hurt the business model of the RecSys.
- Privacy-preserving mechanisms, such as encryption and garbled circuits, make it very difficult to detect Sybil attacks, where an attacker injects fake profiles into the system and then it can (1) try to infer private information based on the outputs to these fake profiles (2) and mount robustness attacks. Canny [9] used zero-knowledge proof technique to fight against ill-formed profiles (i.e. ratings set beyond $\{0, \dots, 5\}$), but it is not effective

against Sybil attacks. With respect to the robustness attacks in reality, the forged rating vectors are always well-formed (but the rating values in these forged rating vectors follow maliciously defined distributions), otherwise the RecSys can easily identify the ill-formed ones in plaintext. To detect and prevent robustness attacks, special detection algorithms need to be executed on the input rating vectors in the privacy-preserving solutions.

When training a recommender model, it is unnecessary to always take into the ratings from all possible users. Amatriain et al. [2] introduced recommender system based on expert opinions, and showed that the recommendation accuracy can be reasonably good even if a target user's data is not used in training the model. Following this concept, Ahn and Amatriain [1] proposed a privacy-preserving distributed recommender system, and similar concept has been adopted in [26,29]. The solution from [29] is very interesting because it leads to very efficient solutions. We briefly summarize it below.

- In the model training stage, suppose the expert data set consists of rating vectors $\mathbf{R}_t$ ($1 \leq t \leq N$). The model parameters are denoted as $\Theta = \{\mathbf{A}, \mathbf{Q}, \mathbf{b}_t^*, \mathbf{b}_j^*\}$, where $\mathbf{A}$ and $\mathbf{Q} = \{\mathbf{q}_1, \dots, \mathbf{q}_M\}$ are two independent item feature spaces, μ is the global rating average, $b_t = \bar{r}_t - \mu$ and $b_j = \bar{r}_j - \mu$ are the average rating for user t and item j respectively, $\mathbf{b}_t^* = \{b_t^*\}_{t=1}^N$, $\mathbf{b}_j^* = \{b_j^*\}_{j=1}^M$ are the user and item bias vectors. Suppose user t has not rated item j , his preference is formulated as follows.

$$\hat{r}_{t,j} = \mu + b_t + b_j + b_t^* + b_j^* + (\mathbf{R}_t \mathbf{A}) \mathbf{q}_j^T \quad (2)$$

Similar to other solutions, SGD can be used to learn the parameter Θ .

- In the prediction computing stage, suppose the user i is not in the expert dataset and has the rating vector $\mathbf{R}_i$ and rating average $\bar{\mathbf{R}}_i$, the prediction for rating $r_{i,j}$ is computed as follows

$$\hat{r}_{i,j} = \bar{\mathbf{R}}_i + b_j + \frac{\sum_{t=1}^N b_t^*}{N} + b_j^* + (\mathbf{R}_i \mathbf{A}) \mathbf{q}_j^T \quad (3)$$

According to their experimental results, the accuracy of the predictions are almost the same to the state-of-the-art recommender systems even though user i is not required to be involved in the model training stage. As such, Wang et al. [29] further proposed an efficient privacy-preserving protocol based on Paillier encryption scheme, so that the prediction (i.e. Equation (3)) can be computed in the encrypted form. Unfortunately, their solution allows a malicious user i or several of such users to straightforwardly recover $b_j + \frac{\sum_{t=1}^N b_t^*}{N} + b_j^*$, $\mathbf{A} \mathbf{q}_j^T$ ($1 \leq j \leq M$), which are functionally equivalent to the model parameters Θ , by solving some simple linear equations. This attack poses severe threats against the *recommendation as a service* objective and the privacy of the RecSys, claimed in [29].

3.3 Examining the DP-based Solutions

While cryptographic solutions might provide provable security for the computation, they do not consider the information leakages from the legitimate outputs. In particular, the inference against an honest user or a group of honest users might be very severe when the attacker has effectively controlled part of the population (e.g. by launching Sybil attacks). Following the seminal work of McSherry and Mironov [19], researchers have tried to apply the differential privacy concept to the prevent information leakages from recommender outputs, e.g. [4,12,13].

One of the main issues with DP-based approach is how to set the privacy parameter ϵ . Specific to recommender systems, it is unrealistic to predefine a privacy budget, because the recommender algorithm (i.e. model training stage) will be executed hundreds, thousands or more times. With respect to the sequential composition theorem, the privacy guarantee becomes $N \cdot \epsilon$ after N executions of the recommender algorithm. In this case, to maintain a meaningful level of privacy protection, the privacy parameter ϵ in every execution needs to be so small such that the recommendation accuracy will be totally destroyed. Besides, most DP-solutions assume a trusted curator (e.g. RecSys), which means there is no privacy against this party. In other solutions (e.g. local differential privacy[25]) no trusted curator is required but it will severely interfere with robustness attack detection operations. For example, a privacy-aware user who prefers a higher level of privacy protection might be prone to be classified as malicious due to the extensive perturbation of his rating vector.

4 Modular Solution Constructions

In this section, we present modular solutions which secure both the model training and prediction computing stages. We first introduce a privacy-by-design concept to minimize information leakages from the outputs, and then describe two types of constructions. In one type of construction, the RecSys trains its recommender model without relying on privacy-aware users' data, while in the other the RecSys needs privacy-aware users' data to train the model so that these users can receive meaningful recommendations. For notation purpose, we refer to them as *Expert-based Solution* and *Self-based Solution* respectively.

Note that for the constructions, we leave the detailed description of privacy-preserving protocols for the prediction computing stage to Section 5 and 6.

4.1 Privacy by Design Concept

In Section 3.2, we have shown that the legitimate outputs in the solutions from [21,29] contain a lot of unnecessary information and can leak the recommender model to a small group of malicious users. To avoid such problems, we enforce the *privacy by design* concept by restricting the output to any user to be the unrated items whose predictions are above a threshold in the proposed prediction

computing stage. This significantly reduces the leakage of MOD_{params} to the user and also allow more efficient protocol design. In reality, the predictions to many items can be quite close so that it is very subtle to only return Top-k (say $k=20$) items. For example, for MovieLens 1M Dataset with 1 million ratings from 6000 users on 4000 movies⁴, the distribution of predictions is shown in Figure 3, where the horizontal Axis stands for the prediction value and vertical Axis stands for the number of predictions that possess the value. Note that all predicted ratings have been rounded to have one decimal place. Intuitively, as an example, it makes more sense to return the unrated items whose ratings are 4.9 or 5. Put it another way, we only need to return the items whose predicted ratings fall into the set $\{V_1 = 5, V_2 = 4.9\}$.

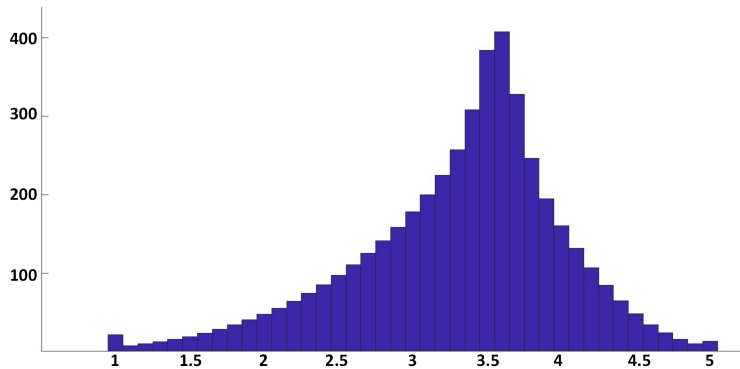


Fig. 3. Prediction Distribution

4.2 Privacy-preserving and Robust Expert-based Solution

In this solution, we adopt the recommender algorithm [29], which has the nice property that the privacy-aware user Alice does not need to share her rating vector with the RecSys to train the recommender model and the process of model training is very simple. Note that in some other expert-based recommender systems, Alice’s data may not be needed to train the model but the process of model training will be much more complex (i.e. often retraining the recommender model is required before being able to generate recommendations for Alice).

1. In the solution, the *model training* stage is very straightforward. Given an expert dataset, the RecSys can first run any robustness attack detection algorithm RobDet to figure out the outliers or even malicious profiles. Then, the RecSys can learn the model parameters $\Theta = \{\mathbf{A}, \mathbf{Q}, \mathbf{b}_i^*, \mathbf{b}_j^*\}$ from the expert dataset, which is publicly available to the RecSys. More information can be seen from Section 3.2.

⁴ <https://grouplens.org/datasets/movielens/1m/>

2. Let's assume that Alice is labelled as user i in the privacy-aware user group, the *prediction computing* stage consists of the following steps.
 - (a) User i generates a public/private key pair (pk_i, sk_i) for an SWHE scheme, and shares the public key pk_i with RecSys.
 - (b) User i sends $\llbracket \mathbf{R}_i \rrbracket_{pk_i}$ and $\llbracket \mathbf{R}_i \rrbracket_{pk_i}$ to the RecSys, which may require the user to prove that the encrypted $\mathbf{R}_i$ is well formed similar to what has been done in [9].
 - (c) If everything is ok, the RecSys can predict user i 's preference on item j as

$$\llbracket \hat{r}_{i,j} \rrbracket_{pk_i} = \llbracket \mathbf{R}_i \rrbracket_{pk_i} \oplus b_j \oplus \frac{\sum_{t=1}^N b_t^*}{N} + b_j^* \oplus \llbracket \mathbf{R}_i \rrbracket_{pk_i} \mathbf{A} \mathbf{q}_j^T \quad (4)$$

- (d) If there is no proxy, user i and the RecSys run the protocol from Section 5 to generate recommendations for user i . Otherwise, they run the protocol from Section 6.

4.3 Privacy-preserving and Robust Self-based Solution

In this solution, we build on top of the privacy-preserving solutions from [21] and [16].

1. In the model training stage, we need to augment existing privacy-preserving protocols for the model training stage, e.g. those from [21] and [16], to enable privacy-preserving robustness attack detection.
 - In case of [21], we need to devise a larger garbled circuit, which first evaluates RobDet and then chooses the unsuspicious inputs to proceed with the matrix factorization procedure.
 - In case of [16], we need to devise a cryptographic protocol that can evaluate RobDet algorithm on the same encrypted inputs to those used in the HE-based matrix factorization algorithm.

A seamless augmentation will depend on the specific robustness attack detection algorithms, so that in this paper we skip the details, which can be an interesting future though.

2. At the end of the privacy-preserving matrix factorization, either from [21] or [16], the RecSys will possess $\llbracket \mathbf{u}_i \rrbracket_{pk_c}$ ($1 \leq i \leq N$) and $\llbracket \mathbf{v}_j \rrbracket_{pk_c}$ ($1 \leq j \leq M$), where (PK_c, SK_c) is an SWHE public/private key pair from the CSP (or Proxy in our system structure). The participants (i.e. user i , RecSys, and Proxy) then perform the following steps.
 - (a) The RecSys computes user i 's preference on item j as $\llbracket \hat{r}_{i,j} \rrbracket_{pk_c} = \llbracket \mathbf{u}_i \rrbracket_{pk_c} \otimes \llbracket \mathbf{v}_j \rrbracket_{pk_c}$ for every $1 \leq j \leq M$.
 - (b) For every $1 \leq j \leq M$, the RecSys selects a random number r_j , then computes and sends $\llbracket \hat{r}_{i,j} \oplus r_j \rrbracket_{pk_c}$ to the proxy. User i generates a Paillier public/private key pair (PK_i, SK_i) and send PK_i to the Proxy.
 - (c) The Proxy decrypts $\llbracket \hat{r}_{i,j} \oplus r_j \rrbracket_{pk_c}$ and re-encrypts the plaintext to obtain $\llbracket \hat{r}_{i,j} \oplus r_j \rrbracket_{pk_i}$ ($1 \leq j \leq M$).
 - (d) For every $1 \leq j \leq M$, the RecSys removes r_j from $\llbracket \hat{r}_{i,j} \oplus r_j \rrbracket_{pk_i}$ to obtain $\llbracket \hat{r}_{i,j} \rrbracket_{pk_i}$.

3. If there is no proxy, user i and the RecSys run the protocol from Section 5 to generate recommendations for user i . Otherwise, they run the protocol from Section 6.

It is clear that the model training stage of our expert-based solution satisfies all our robustness and privacy expectations, while the privacy analysis depends on the protocols from Section 5 and 6 because the existing steps do not leak information due to the encrypted operations and randomization. For the self-based solution, we can guarantee the same level of privacy and robustness protection, although it will apparently be less efficient than the previous expert-based one.

5 Privacy-preserving Prediction Computing

In this section, we describe a privacy-preserving protocol for user i to learn the unrated items whose predictions fall into a set $\{V_1, \dots, V_T\}$, without relying on a proxy. Here T will be a small integer, which may be 2 or 3 in practice referring to the example in the previous section. Observing that privacy-preserving protocols for the model training stage often output integer predictions (in encrypted form), because they need to scale the intermediary computation results in order to be compatible with the cryptographic tools such as homomorphic encryption algorithms. Therefore, we assume the RecSys possesses the encrypted predictions $\llbracket x_j \cdot \theta + y_j \rrbracket_{pk_i}$ for every $1 \leq j \leq M$ at the end of the privacy-preserving model training stage. We explicitly present the ratings according to a unit θ , because in our protocol the recommendations will only be based on the x_j part and the y_j part is rounding off.

5.1 Description of the Proposed Protocol

At the beginning of the prediction computing stage, we suppose user i possesses two public/private key pairs: one is (PK_i, SK_i) for the Paillier scheme which has been setup in Section 4.2 and 4.3 while the other is new key pair (PK'_i, SK'_i) for a SWHE encryption scheme [18]. The public keys PK_i, PK'_i are shared with the RecSys. As shown in Figure 4, the protocol runs in two phases where λ is the security parameter.

In the *reduction* phase, the RecSys and user i round off the y_j part in the encrypted predictions. Specifically, for every $1 \leq j \leq M$, the following operations will be carried out.

1. The RecSys first randomizes x_j and y_j to generates Δ_j for user i .
2. Then, user i obtains the randomized prediction value α_j through decryption and then computes β_j , which is the randomized x_j in an approximation form with $\epsilon_j \in \{0, 1\}$. Finally, user i encrypts β_j under his own SWHE public key if item j is unrated, and encrypts a random value otherwise.
3. After receiving Γ_j , the RecSys homomorphically removes the randomization noise r_{j1} to obtain Φ_j , which is a ciphertext for $x_j + \epsilon_j$ if item j is unrated and a ciphertext for a random value otherwise.

	User i Paillier : (PK_i, SK_i) SWHE : (PK'_i, SK'_i)	RecSys $\llbracket x_j \cdot \theta + y_j \rrbracket_{pk_i}$ $(1 \leq j \leq M)$
Reduction	for each $1 \leq j \leq M$ Δ_j $\longleftarrow$ $\alpha_j = \mathbf{Dec}(\Delta_j, SK_i)$ $= (x_j + r_{j1}) \cdot \theta + y_j + r_{j2}$ $\beta_j = \frac{\alpha_j - (\alpha_j \bmod \theta)}{\theta}$ $= x_j + r_{j1} + \epsilon_j$ If item j is unrated : $\Gamma_j = \mathbf{Enc}(\beta_j, PK'_i)$ Otherwise : $r_{j3} \xleftarrow{\\$} \{0, 1\}^\lambda$ $\Gamma_j = \mathbf{Enc}(r_{j3}, PK'_i)$ Γ_j $\longrightarrow$	$r_{j1} \xleftarrow{\$} \{0, 1\}^\lambda$ $r_{j2} \xleftarrow{\$} [0, \theta)$ $\Delta_j = \llbracket x_j \cdot \theta + y_j \rrbracket_{pk_i} \oplus (r_{j1} \cdot \theta + r_{j2})$ $= \llbracket (x_j + r_{j1}) \cdot \theta + y_j + r_{j2} \rrbracket_{pk_i}$
Evaluation	$\Psi_j (1 \leq j \leq M)$ $\longleftarrow$ $\{j \mathbf{Dec}(\Psi_j, SK'_i) \neq 0\}$	$\Phi_j = \Gamma_j \ominus r_{j1}$ $\Omega_j = \prod_{x=1}^T (\Phi_j \ominus V_x)$ $\Psi_j = \mathbf{RAND}(\Omega_j)$

Fig. 4. Learning Membership in $\{V_1, \dots, V_T\}$ without a Proxy

In the *Evaluation* phase, for every $1 \leq j \leq M$, the RecSys computes Ω_j through T homomorphic subtractions and $T - 1$ multiplications, which is a ciphertext for 0 if the plaintext corresponding to Φ_j falls into $\{V_1, \dots, V_T\}$ and a ciphertext for a non-zero value otherwise. In order to hide the non-zero values, the RecSys randomize Ω_j via the RAND function, e.g. homomorphically multiplying a random number, to obtain Ψ_j , which can be decrypted by user i to learn the index of recommended items.

5.2 Security and Performance Analysis

The operations in the protocol are done with encrypted data and randomization has been applied to the predictions revealed to user i . As such the protocol only reveals the desired items to user i while leaks nothing to the RecSys.

For Paillier, we set the size of N to be 2048, and for SWHE we use Microsoft SEAL library. We select the ciphertext modulus $q = 2^{226} - 2^{26} + 1$, the polynomial modulus $p(x) = x^{8192} + 1$. Using Chinese Remainder Theorem, we select two 40-bit primes to represent the plaintext space of 2^{80} . The primes are 1099511922689 and 1099512004609. By packing 8192 plaintexts into one ciphertext, we can process 8192 multiplications in one homomorphic multiplication. Based on an Intel(R)

Core(TM) i7-5600U CPU 2.60GHz, 8GB RAM, the timing is summarized in Table 1.

Paillier.Enc	Paillier.Dec	Paillier. $\oplus$	SWHE.Enc
31.30 ms	12.88 ms	8.50 μ s	52.43 ms
SWHE.Dec	SWHE. $\otimes$	partial SWHE. $\otimes$	SWHE. $\oplus$
39.63 ms	207.76 ms	70.28 ms	742 μ s

Table 1. Costs for SWHE and Paillier

The number of different cryptographic operations for the proposed protocol are summarized in Table 2. In the last column, we estimate the real-world running time based on the aforementioned benchmarking results, where $M = 4000$ by assuming the MovieLens 1M Dataset and $T = 2$. Note that this dataset has been used in Section 4.1.

	Paillier.Dec	Paillier. $\oplus$	SWHE.Enc	SWHE.Dec	SWHE. $\otimes$	partial SWHE. $\otimes$	SWHE. $\oplus$	Time
User	M		M	M				420 s
RecSys		M		M	$M(T - 1)$		$M(T + 1)$	998 s

Table 2. Computational Complexities

With respect to the MovieLens 1M Dataset, we consider the standard case where user i and the RecSys interactively rank the predictions and the RecSys returns the top-ranked items. In order to rank, user i and RecSys need to perform a comparison for two predictions for the RecSys to learn the order of them. Based on the same computer as above, for a comparison with the protocol from [7], the computation time for user i and the Recsys is 175.88 ms and 184.60 ms respectively. Suppose we adopt a standard sorting algorithm to realise the ranking, and the average computation time for the user and RecSys will be 8442.24 s and 8860.80 s, respectively. The time delay due to the communication is about 4800 s, by assuming each computation takes up to 100 ms as in [7]. It is clear that our protocol is much more efficient.

6 Privacy-preserving Prediction Computing with Proxy

In this section, we describe the protocol that relies on a proxy, and also provide corresponding analysis.

6.1 Description of the Proposed Protocol

To enable the new protocol, we make use of a key-homomorphic pseudorandom function Prf [3]. Given $\text{Prf}(k_1, m)$ and $\text{Prf}(k_2, m)$, anybody can compute $\text{Prf}(k_1 + k_2, m) = \text{Prf}(k_1, m) \oplus \text{Prf}(k_2, m)$. We describe the two phases in Figures 5 and 6, respectively. As before, λ is the security parameter.

	User i	RecSys
	$\text{Paillier} : (PK_i, SK_i)$ $\text{SWHE} : (PK'_i, SK'_i)$	$\llbracket x_j \cdot \theta + y_j \rrbracket_{pk_i}$ ($1 \leq j \leq M$)
Reduction		$r_{j1} \xleftarrow{\$} \{0, 1\}^\lambda$ $r_{j2} \xleftarrow{\$} [0, \theta)$ $\Delta_j = \llbracket x_j \cdot \theta + y_j \rrbracket_{pk_i} \ominus (r_{j1} \cdot \theta - r_{j2})$ $= \llbracket (x_j - r_{j1}) \cdot \theta + y_j + r_{j2} \rrbracket_{pk_i}$
for each $1 \leq j \leq M$	$\xleftarrow{\Delta_j}$ $\alpha_j = \text{Dec}(\Delta_j, SK_i)$ $= (x_j - r_{j1}) \cdot \theta + y_j + r_{j2}$ $\beta_j = \frac{\alpha_j - (\alpha_j \bmod \theta)}{\theta}$ $= x_j - r_{j1} + \epsilon_j$ $r_{j3} \xleftarrow{\$} \{0, 1\}^{2\lambda}$ If item unrated : $\gamma_j = r_{j3} + \beta_j$ $= x_j - r_{j1} + \epsilon_j + r_{j3}$ If item rated : $\gamma_j = r_{j3}$	
	$\xrightarrow{\gamma_j}$	$\gamma_j = \gamma_j + r_{j1}$

Fig. 5. Reduction Phase

Similar to the case shown in Figure 4, in the *reduction* phase, the RecSys and user i interactively round off the y_j part in the predictions for every j . The main difference (and simplification) is that, at the end of the protocol, the RecSys possesses $\gamma_j = x_j + \epsilon_j + r_{j3}$ if item j has been rated and $\gamma_j = r_{j3} + r_{j1}$ otherwise, while user i possesses the random number r_{j3} .

User i	Proxy	RecSys
$r_{j3} (1 \leq j \leq M)$		$\gamma_j (1 \leq j \leq M)$
$R_j, \text{PM}_j (1 \leq j \leq M)$ PM, H $K_j (1 \leq j \leq M) \leftarrow \{0, 1\}^{3\lambda}$ $\mathcal{T}_j = \text{Prf}(K_j - r_{j3}, R_j)$ ($1 \leq j \leq M$)		$R_j, \text{PM}_j (1 \leq j \leq M)$ PM, H $\Theta_j = \text{Prf}(\gamma_j, R_j)$ ($1 \leq j \leq M$)
	$\xrightarrow{\text{PM}\{\{\mathcal{T}_1, \dots, \mathcal{T}_M\}\}}$	$\xleftarrow{\text{PM}\{\{\Theta_1, \dots, \Theta_M\}\}}$
$\Omega_{x,j} = \text{H}(\text{Prf}(K_j + V_x, R_j))$ ($1 \leq x \leq T, 1 \leq j \leq M$) $\Omega_{*,j} = \text{PM}_j\{(\Omega_{1,j}, \dots, \Omega_{T,j})\}$	$\xrightarrow{\text{PM}\{(\Omega_{*,1}, \dots, \Omega_{*,M})\}}$	$\mathbb{S}$
	$\xleftarrow{\mathbb{S}}$	

Fig. 6. Evaluation Phase (w.r.t. $\{V_1, \dots, V_T\}$)

The evaluation phase, shown in Figure 6, proceeds as follows.

1. User i first establishes M random messages $R_j (1 \leq j \leq M)$, random permutation functions PM and $\text{PM}_j (1 \leq j \leq M)$, and a hash function H with the RecSys. Given a vector of M elements, PM randomly permutes the order of the elements. Similarly, given a vector of T elements, PM_j randomly permutes the order of the elements.
2. User i chooses M random keys $K_j (1 \leq j \leq M)$ for Prf and evaluates Prf for R_j with the key $K_j - r_{j3}$ to obtain Υ_j , for every $1 \leq j \leq M$. At the same time, the RecSys evaluates Prf for R_j with the key γ_j to obtain Θ_j , for every $1 \leq j \leq M$.
3. After receiving the permuted values from user i and the RecSys, the proxy computes

$$\Xi = \text{PM}\{(\Upsilon_1, \dots, \Upsilon_M)\} \oplus \text{PM}\{(\Theta_1, \dots, \Theta_M)\},$$

where $\oplus$ is performed element wise. It is easy to check that if the item j is unrated

$$\Upsilon_j \oplus \Theta_j = \text{Prf}(K_j + x_j + \epsilon_j, R_j), \text{ and otherwise}$$

$$\Upsilon_j \oplus \Theta_j = \text{Prf}(K_j + r_{j1}, R_j).$$

4. User i first computes $\Omega_{x,j} = \text{H}(\text{Prf}(K_j + V_x, R_j))$ for every $1 \leq x \leq T, 1 \leq j \leq M$, and then computes a randomized check value vector $\Omega_{*,j} = \text{PM}_j\{(\Omega_{1,j}, \dots, \Omega_{T,j})\}$ for every item j . It permutes a vector, formed by individual check value vectors of all items, and sends the result $\text{PM}\{(\Omega_{*,1}, \dots, \Omega_{*,M})\}$ to the RecSys.
5. After receiving $\text{PM}\{(\Omega_{*,1}, \dots, \Omega_{*,M})\}$ from the user, the proxy can compute $\mathbb{S}$, which is a new set generated based on Ξ : for every element in Ξ , if its hash value with respect to H appears in the corresponding element in $\text{PM}\{\Omega_{*,j} (1 \leq j \leq M)\}$ then the corresponding element in $\mathbb{S}$ is set to be 1 otherwise it is set to be 0.
6. With $\mathbb{S}$ and PM , user i can identify the unrated items whose approximated predictions fall into the set $\{V_1, \dots, V_T\}$.

6.2 Security and Performance Analysis

With encryption, the *reduction* phase leaks no information to either party. In the *evaluation* phase, the Recsys does not learn anything because it receives nothing from others, while user i only learns which items are recommended. Regarding the information leakage to the proxy, we only need to discuss an item j for any $1 \leq j \leq M$, because different K and R are used for different items. For any j , due to the fact that K_j, r_{j3}, r_{j1} are chosen independently and at random, the Υ_j, Θ_j are random values in the view of the proxy. With H being modelled as a random oracle, $\Omega_{*,j}$ leaks no information if item j has been rated, and it only tells whether $\text{Prf}(K_j + x_j + \epsilon_j, R_j)$ is a match and nothing else. The random permutations PM hides which items have been recommended to the user i , while $\text{PM}_j (1 \leq j \leq M)$ hide the predicted rating values for the recommended items. With respect to the security model from Section 2.1, the solution leaks the number of recommended items to the proxy, while in the security model it is

required that there should be no leakage. To reduce the leakage, we can replace Step 4-6 with a privacy-preserving set interaction protocol. We leave a detailed investigation of this issue as a future work.

We summarize the asymptotic complexity in Table 3. Based on the reference codes by the authors of [3]⁵, the `Prf.Evaluate` and `Prf.Hadd` takes about 1.04 ms and 10 μ s. W.r.t. the MovieLens 1M Dataset and $T = 2$, we compute the real-world running time and put it in the last column of Table 3. It is clear that the existence of Proxy greatly improves the efficiency without seriously downgrading the privacy guarantee.

	Paillier.Dec	Paillier. $\oplus$	Prf.Evaluate	Prf.Hadd	Time
User	M		$M(1 + T)$		63.52 s
RecSys		M	M		4.16 s
Proxy				M	40 ms

Table 3. Computational Complexities

7 Conclusion

In this paper, we have demonstrated how to construct privacy-preserving collaborative filtering recommenders by separately addressing the privacy issues in the model training and prediction computation stages. We argued that the expert-based approach (e.g. [29]) provides more scalable solution to the model training stage, while the efficiency of existing cryptographic solutions (e.g. [21] and [16]) remains as a challenge particularly with the need to support robustness attack detection. By leveraging homomorphic encryption and key-homomorphic pseudorandom functions, we show that the proposed privacy-preserving prediction computing protocols are much more efficient than standard solutions. The current paper leaves several interesting research questions. One is to investigate the performances of cryptographic solutions when they are extended to support robustness attack detection and also improve their efficiency. Another research question is to formally study the privacy advantage of the *privacy by design* approach in providing recommendations to end users, and potentially link it to differential privacy. Yet another research question is to investigate the performances (e.g. recommendation accuracy) of the two privacy-preserving protocols for the prediction computing stage based on other widely-used datasets such as Netflix.

References

1. J. Ahn and X. Amatriain. Towards fully distributed and privacy-preserving recommendations via expert collaborative filtering and restful linked data. In *2010*

⁵ <https://github.com/cpeikert/Lol/tree/master/lol-apps>

- IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010*, pages 66–73, 2010.
2. X. Amatriain, N. Lathia, J. M. Pujol, H. Kwak, and N. Oliver. The wisdom of the few: A collaborative filtering approach based on expert opinions from the web. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 532–539, 2009.
 3. A. Banerjee and C. Peikert. New and improved key-homomorphic pseudorandom functions. In *Advances in Cryptology - CRYPTO 2014*, pages 353–370, 2014.
 4. A. Berlioz, A. Friedman, M. A. Kaafar, R. Boreli, and S. Berkovsky. Applying differential privacy to matrix factorization. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 107–114, 2015.
 5. M. Beye, A. Jeckmans, Z. Erkin, Q. Tang, P. Hartel, and I. Lagendijk. *Social Media Retrieval*, chapter Privacy in Recommender systems, pages 263–281. Springer, 2013.
 6. J. W. Bos, K. E. Lauter, J. Loftus, and M. Naehrig. Improved security for a ring-based fully homomorphic encryption scheme. In *Cryptography and Coding - 14th IMA International Conference*, pages 45–64, 2013.
 7. R. Bost, R. A. Popa, S. Tu, and S. Goldwasser. Machine learning classification over encrypted data. In *22nd Annual Network and Distributed System Security Symposium, NDSS 2015*, 2015.
 8. J. Canny. Collaborative filtering with privacy via factor analysis. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 238–245, 2002.
 9. J. F. Canny. Collaborative filtering with privacy. In *IEEE Symposium on Security and Privacy*, pages 45–57, 2002.
 10. Z. Cheng and N. Hurley. Trading robustness for privacy in decentralized recommender systems. In *Proceedings of the Twenty-First Conference on Innovative Applications of Artificial Intelligence*, pages 3–15, 2009.
 11. N. Dowlin, R. Gilad-Bachrach, K. Laine, K. E. Lauter, M. Naehrig, and J. Wernsing. Manual for using homomorphic encryption for bioinformatics. *Proceedings of the IEEE*, 105(3):552–567, 2017.
 12. C. Dwork. Differential privacy. In M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, editors, *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006*, volume 4052 of *LNCs*, pages 1–12. Springer, 2006.
 13. C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography, Third Theory of Cryptography Conference*, pages 265–284, 2006.
 14. A. Friedman, B. P. Knijnenburg, K. Vanhecke, L. Martens, and S. Berkovsky. *Recommender Systems Handbook*, chapter Privacy Aspects of Recommender Systems, pages 649–688. Springer, 2015.
 15. A. Jeckmans, A. Peter, and P. H. Hartel. Efficient privacy-enhanced familiarity-based recommender system. In *Computer Security - ESORICS 2013*, pages 400–417, 2013.
 16. S. Kim, J. Kim, D. Koo, Y. Kim, H. Yoon, and J. Shin. Efficient privacy-preserving matrix factorization via fully homomorphic encryption: Extended abstract. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, pages 617–628, 2016.
 17. S. K. Lam and J. Riedl. Shilling recommender systems for fun and profit. In *Proceedings of the 13th International Conference on World Wide Web*, pages 393–402, 2004.

18. Microsoft SEAL Library. <https://sealcrypto.codeplex.com/>, 2016.
19. F. McSherry and I. Mironov. Differentially private recommender systems: building privacy into the Netflix prize contenders. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 627–636, 2009.
20. B. Mobasher, R. Burke, and J. J. Sandvig. Model-based collaborative filtering as a defense against profile injection attacks. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, pages 1388–1393, 2006.
21. V. Nikolaenko, S. Ioannidis, U. Weinsberg, M. Joye, N. Taft, and D. Boneh. Privacy-preserving matrix factorization. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, pages 801–812, 2013.
22. P. Paillier. Public-key cryptosystems based on composite degree residuosity classes. In *Advances in Cryptology - EUROCRYPT 1999*, pages 223–238, 1999.
23. H. Polat and W. Du. Privacy-preserving collaborative filtering using randomized perturbation techniques. In *Proceedings of ICDM 2003*, pages 625–628, 2003.
24. G. Shani and A. Gunawardana. Evaluating recommendation systems. In *Recommender Systems Handbook*, pages 257–297. Springer, 2011.
25. Y. Shen and H. Jin. Epicrec: Towards practical differentially private framework for personalized recommendation. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 180–191, 2016.
26. Q. Tang and H. Wang. Privacy-preserving hybrid recommender system. In *The Fifth International Workshop on Security in Cloud Computing (SCC)*, pages 59–66. Springer, 2017.
27. Q. Tang and J. Wang. Privacy-preserving context-aware recommender systems: Analysis and new solutions. In *Computer Security - ESORICS 2015*, pages 101–119, 2015.
28. F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart. Stealing machine learning models via prediction apis. In *Proceedings of the 25th USENIX Conference on Security Symposium*, pages 601–618, 2016.
29. J. Wang, A. Arriaga, Q. Tang, and P. Y. A. Ryan. Novel collaborative filtering recommender friendly to privacy protection, 2019.
30. U. Weinsberg, S. Bhagat, S. Ioannidis, and N. Taft. Blurme: inferring and obfuscating user gender based on ratings. In *Sixth ACM Conference on Recommender Systems*, pages 195–202, 2012.
31. S. Zhang, J. Ford, and F. Makedon. Deriving private information from randomly perturbed ratings. In *Proceedings of the Sixth SIAM International Conference on Data Mining*, pages 59–69. SIAM, 2006.
32. S. Zhang, L. Yao, and A. Sun. Deep learning based recommender system: A survey and new perspectives. <https://arxiv.org/abs/1707.07435>, 2017.