Multimed Tools Appl (2016) 75:14927–14957
DOI 10.1007/s11042-014-2271-0

Collaborative privacy framework for minimizing privacy risks in an IPTV social recommender service

Ahmed M. Elmisery · Seungmin Rho · Dmitri Botvich

Received: 28 April 2014 / Revised: 4 September 2014 / Accepted: 9 September 2014 /
Published online: 21 September 2014

Abstract In our connected world, recommender systems have become widely known for their ability to provide expert and personalized referrals to end-users in different domains. The rapid growth of social networks has given a rise to a new kind of systems, which have been termed "social recommender service". In this context, a software as a service recommender system can be utilized to extract a set of suitable referrals for certain users based on the data collected from the personal profiles of other end-users within a social structure. However, preserving end-users privacy in social recommender services is a very challenging problem that might prevent privacy concerned users from releasing their own profiles' data or to be forced to release an erroneous data. Thus, both cases can detain the accuracy of extracted referrals. So in order to gain accurate referrals, the social recommender service should have the ability to preserve the privacy of end-users registered in their system. In this paper, we present a middleware that runs on the end-users' side in order to conceal their profiles data when being released for the recommendation purposes. The computation of recommendation proceeds over this concealed data. The proposed middleware is equipped with a distributed data collection protocol along with two stage concealment process to give the end-users complete control over the privacy of their profiles. We will present an IPTV network scenario along with the proposed middleware. A number of different experiments were performed on real data which was concealed using our two stage concealment process to evaluate the achieved privacy and accuracy of the extracted referrals. As supported by the experiments, the proposed framework maintains the recommendations accuracy with a reasonable privacy level.

Keywords Privacy · Clustering · IPTV network · Recommendation systems

A. M. Elmisery · D. Botvich
TSSG, Waterford Institute of Technology-WIT-Co, Waterford, Ireland

A. M. Elmisery
e-mail: ahmedmohmed2001@gmail.com

D. Botvich
e-mail: dbotvich@tssg.org

S. Rho (✉)
Department of Multimedia, Sungkyul University, Anyang-si, South Korea
e-mail: smrho@sungkyul.edu

1 Introduction

Internet Protocol Television (IPTV) is a video service providing IP broadcasts and video on demand (VOD) over a broadband IP content delivery network (CDN) specialized in video services. The IPTV user has an access to myriads of video content spanning IP Broadcast and VOD [31]. In this context, it is difficult for the end-users to find a content that matches their preferences from the huge amount of video contents that they are confronting on while using the IPTV system [28]. Moreover, gathering all the information to make a well-grounded decision to buy or watch a specific content is a very time consuming process. In order to attract and satisfy these users, the IPTV service providers employ recommendation systems to increase their revenues and offer added values to their patrons. Recommendation System is a promising personalization system especially for the IPTV services where it offers referrals to the end-users by capturing their preferences using either explicit or implicit methodologies to create preferences profiles based on their consumption history, behaviour, purchased transactions and demographic information. In the context of this paper, a profile is a list that comprises the video contents that the user has watched or purchased combined with the meta-data extracted from the content provider regarding this content (i.e. genres, directors, actors and so on) and the ratings that the user gave to these contents. In such a way, recommender systems can assist the end-users in quickly making the proper decision, and save their time and money. Therefore, the recommender systems are usually referred to as the experts and they have been used in crucial fields like healthcare services, financial investments, and e-learning. It is believed that the recommender systems can substitute experts, not only because employing automatic recommender systems is cheaper than hiring an expert [47], but also because the generated referrals can outperform the advice of an expert. Social recommendation systems are usually served using collaborative filtering (CF) algorithms, which are a popularly used technical approach to automate the word-of-mouth process; it is based on the hypothesis that people with similar tastes prefer the same items. The recommendation using CF technique involves a main entity that collects users' profiles to find a set of users similar to the user receiving the recommendation which will be denoted as the target user. Then after, it executes the CF algorithms to suggest to him/her the contents that have been rated high in the past by the other users. From what we have mentioned before, we can infer that recommender systems are automatic systems and they can generate personalized results that the users were not aware of. A software-as-a-service recommender system is a new business model [33] which realizes a third-party company offering the functions of recommender systems as a service to a set of registered clients over a service oriented infrastructure. These recommender services host users' data from the various content providers then employ various techniques in flexible and transparent configurations in order to extract referrals. Finally, these referrals are delivered through APIs to the clients. These services can scale to serve multiple content providers, thus, this large providers' base leads to cost reduction in service leasing, in contrast to a situation where in-house recommender systems were deployed and operated by the content providers themselves.

Privacy is a necessity for the recommender services, as the recommendation process requires a detailed view/profile of each user. While In the general case, collecting high quality detailed profiles from users is desirable as the recommendations can be highly beneficial for both of the users and the IPTV service providers, but it is not an easy task as the price is high, likewise: total loss of privacy while generating the recommendations. These profiles can include sensitive information that capture the personal description of a particular user, which represent a serious invasion to the individual privacy as the collected profiles can be used for unsolicited marketing, government surveillance, profiling users, or it can be sold to external parties when the service providers face bankruptcy. However, some users are willing to reveal their whole profiles in order to get accurate referrals but others may be concerned about the privacy implications of disclosing their profiles that can open a door for the misuse of their personal data. However, the current business model for those recommender services is centered around the availability of users' personal data at their side whereas users have to trust that the recommender service providers will not use their data in a malicious way. With the increasing number of cases for privacy breach of personal information, the need has increased for tools or technologies enabling the end-users to have control over their privacy. Currently there are two options for the privacy concerned users when using IPTV recommender system: firstly, they can refuse to enter the information they are uncomfortable about disclosing, which in turn brings the sparse data problem [25] for the recommendation technique since only a subset of items ha been released by each user. Secondly, they may enter fake information, which in turn decreases the accuracy of the generated recommendations and leads to lack of acceptance of the recommendation process in general. As a matter of fact, an actual rating given to an item by a user produces a reasonable explanation and an accurate ranking from a reliable source. Users are more likely willing to give more truthful data if privacy measurements are provided during data collection or if they have assured their privacy will be preserved. Privacy enhancing frameworks have emerged in order to meet the privacy requirements of end-users. These frameworks can be future divided into technological and legislation solutions. The former approach refers to technical methods and tools that are integrated into systems or networks to reduce the collection of accurate personal data. Such methods and tools are referred to as privacy enhancing technologies (PETs). The latter refers to data protection legislation restricting the gathering and usage of private personal data by the data processors in order to define the best practices for the protection of personal information. Four examples for such privacy guidelines are the EU Directives 95/46/EC [9] and 2002/58/EC [7], UK's Data Protection Act and OECD privacy principles [8]. Our view in this paper takes into consideration that recommender services should also take care of the privacy of end-users during the recommendation process because the data collected from IPTV users cover their personal preferences about different contents that they have watched or purchased. Our solution relies on a holistic approach for achieving privacy by developing a privacy enhancing technology that has been designed according to the guidelines of a legal privacy principle that is the OECD's recommendations for protection of personal data. The proposed collaborative privacy framework unifies the legal and technical regulations together in one user-centric privacy framework in order to implement and impose the legal privacy principles for the protection of personal data and enables the end-users to have full control over what is being released or collected from their profiles' data. Due to the frequent complexities that arise when integrating any newly-proposed privacy enhancing frameworks within the current recommender service back-ends, our solution utilizes the user and social sides of the social recommender services as an infrastructure for the proposed collaborative privacy framework.

In this paper, we will present our collaborative privacy framework that has been implemented using an Enhanced Middleware for Collaborative Privacy (*EMCP*). The proposed framework allows the creation of serendipity recommendations without breaching users' privacy. The proposed framework utilizes social coalition to attain better privacy, where implicit groups are created between the end-users to fulfil a specific recommendation process. The users' cooperation is needed not only to protect their privacy but also to make the service run properly. The collaborative privacy framework employs a two stage concealment process to allow the end-users to privately release their data between each other within the coalition and also to attain anonymity during the recommendation process. The two stage concealment process secures the user's rating profile in the untrusted social recommender service with minimum loss of accuracy. The first process in the two stage concealment process is the local concealment process which executes the clustering based obfuscation algorithm (CBO) in order to locally conceal the user's rating profile before releasing it outside of his/her device. The CBO algorithm divides the user rating profile into smaller clusters and then introduces a carefully chosen artificial noise to these clusters in order to retain its statistical content while concealing all private information. The other stage of the two stage concealment process is the global concealment process which is executed remotely at the trusted aggregator within the coalition. The trusted aggregator is the entity that is responsible for collecting the profiles of end-users before releasing them to the social recommender service. The global concealment process executes the random ratings generation (RRG) algorithm in order to alleviate data sparsity problem within the collected profiles by filling the unrated cells in the dataset in such a way to improve recommendation accuracy and increase the attained privacy. The collaborative privacy approach preserves the aggregates in the dataset to maximize the usability of information in order to accurately predicate ratings for items that have not been consumed before by the target-user. In rest of this paper, we will generically refer to news programs, movies and video on demand contents as Items. This paper is organised as follows. In Section 2, the background about this work is presented. Section 3 introduces a scenario for the IPTV content distribution network landing our collaborative privacy framework. The proposed solution based on *EMCP* is introduced in Section 4. The proof of security for the two-stage concealment process is demonstrated in Section 5. In Section 6, the Results from some experiments on the proposed framework are presented. The paper ends with Section 7, which presents the conclusion and future work.

2 Background

2.1 Software-as-a-service recommender system

Some of the IPTV content providers can run the recommender system as an internal service within their distribution network. On the one hand, these providers are required to buy, build, train and maintain their recommender service infrastructure despite exponential costs. Additionally, in order to run this service well, these providers are required to recruit a highly specialized team to tune and handle the ongoing problems that arise when this kind of services runs. On the other hand, obtaining recommender system as a service gives the providers an alternative. Now, they can plug in and subscribe to any social recommender service that was built on a shared infrastructure via the Internet. The popularity of software as services solutions is steadily increasing because they simplify the deployment and reduce the customer acquisition costs. With the social recommender service, the IPTV providers can support many users with a single version of the service. The multitenancy feature of social recommender service allows the IPTV providers to scale as fast and as much as needed without replacing the costly infrastructure or adding a new IT staff. The architecture for the social recommender service can be described as Fig. 1.

The social recommender service is implemented using the service-oriented architecture (SOA), which has flourished in recent years due to the numerous benefits this architecture can offer to businesses of all sizes and types. Here's what's driving the IPTV providers to employ the social recommender service solution:

- High Adoption: Social recommender service solutions are available from any computer or any device. Because most of the users are using the internet to find what they want, social

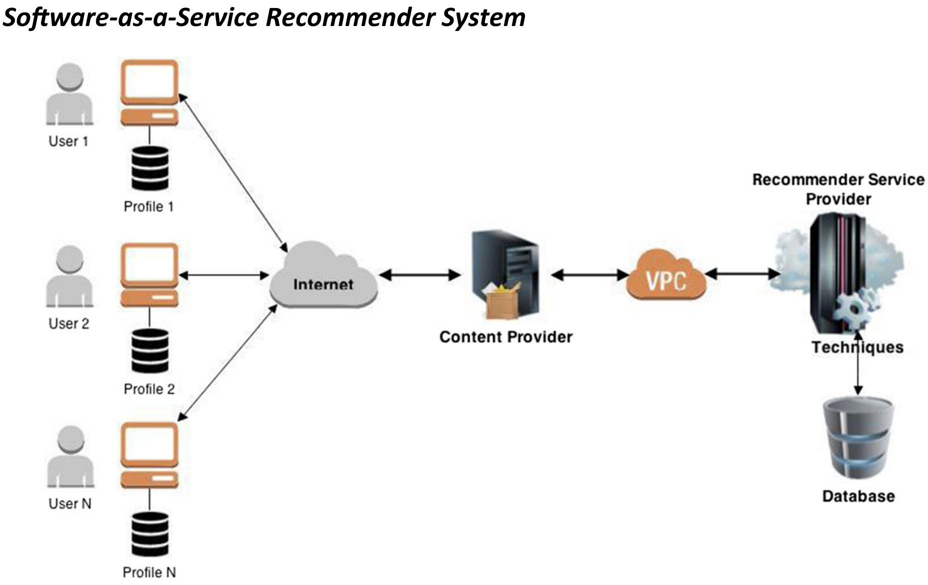


Fig. 1 Architecture of social recommender service

recommender service tends to have a high reliability proportion, with a rapid proficiency curve.

- Lower Initial Costs: Social recommender service solutions are based on the subscription model. There is no need to pay for license fees, which in turn implies lower start-up costs for this model. Having an external service provider manages the IT infrastructure of social recommender service means lower IT costs for hardware, software, and the staff needed at the IPTV content provider side.
- Painless Upgrades: Because an external service provider manages all the updates and upgrades, there are no patches for users to download or install. The external service provider also manages availability, so there is no need for users to upgrade their hardware, buy a new software, or increase the bandwidth as the user base grows.
- Seamless Integration: The providers of social recommender service with a true multitenant architecture can scale indefinitely to meet the customers' demands. Many recommender service providers also offer customization capabilities to meet the specific needs of the content providers. Additionally, many recommender service providers provide a set of APIs to facilitate the integration with the existing ERP or business productivity systems on the IPTV content provider side.

As illustrated before, IPTV content providers can utilize an external social recommender as a service in order to provide referrals for their clients [47]. An external recommender service utilizes a set of basic steps in order to extract referrals from the users' profiles. The actual recommender services may reduce some of these steps while others can perform complex stages within some of them:

1. The IPTV content providers insert the profiles of items and descriptions about these items into the local database of the social recommender service. These profiles contain the main

characteristics of every item. This is an essential step as it allows the recommender service to identify the items that are going to be offered to the users.

2. A profile is created and assigned to each user registered at the IPTV content provider. This profile could be controlled by the user or IPTV content provider. These profiles capture the preferences, ratings, and personal information regarding the system's users. Therefore, the information contained in these profiles is highly sensitive which imposes a responsibility on the IPTV content providers to protect it in order to increase the trustworthiness of their services.
3. Users send a request to the recommender service that includes their preferences. The complexity of this process varies with different recommender types.
4. The recommendation algorithms/techniques are utilized in order to respond to that request, whereas recommenders search their internal database to select items that are appropriate to answer this request. The recommender service returns a set of identifiers for items that might be interesting to the users. These identifiers are linked to items which are offered by the IPTV content providers.
5. During the final phase, users access the recommended items. These generated referrals may be useful to enhance the service and future recommendations, as accessing those items does imply that the recommendations were correct.

This model allows the IPTV content providers to utilize the advanced computing resources, techniques, and expertise of the social recommender service for generating referrals based upon profiles collected from users in their distribution networks. IPTV Content providers act as a mediator for accumulating users' profiles and delivering the recommended contents to their subscribers. As we explained beforehand, the collected profiles contain sensitive information, which raise privacy concerns for the system's users [29]. The main privacy threats of such a model are:

1. How can users make sure that the recommender service provider is trusted enough to handle their profiles' data in the same way they are announced?
2. How can users make sure that the recommender service provider prohibits unauthorized parties from accessing their data?

Although a reliable behaviour from the recommender service provider can be imposed by a contract, however, due to various cases, the legislation alone might not be enough. The rapid development in technology, differences between privacy laws, complex breaches in the infrastructure of the recommender service provider, and finally, the difficulty in detection and prevention of violations in the outsourced data, all of which limit the feasibility of any legislative efforts. The goal of this paper is to propose a collaborative privacy framework to ensure the privacy of the data outsourced to the social recommender service as an input while allowing the extraction of accurate referrals from this data. The proposed framework conceals the outsourced profiles' data in a way which enables the recommender service to execute its desired recommendation technique on such concealed data yielding accurate referrals compared with the ones extracted from the real data.

2.2 Challenges in designing our collaborative privacy framework for IPTV network scenario

Despite the benefits of the utilization of social recommender services in IPTV content distribution networks, new privacy threats exist while making use of these external services especially when combined with the social network side. As bringing various data together to support these services make misuse easier, yet in the absence of adequate safeguards, the frequent use of these services can jeopardize the privacy and autonomy of users. Privacy invasion occurs when individuals are unaware of "behind the scenes" use of their personal data. The simplest form of privacy invasion by social recommender service providers are unsolicited marketing, customer segmentation, and scoring [8]. Data collected from the users is a valuable asset, and it can be sold when providers suffer bankruptcy.

When we started to design our proposed collaborative privacy framework, we discovered that designing such framework for social recommender services is different than designing a privacy enhancing framework for any other kind of services. This is due to the inherented problems that exist within the social recommender service that limit reusing conventional versatile privacy enhancing frameworks that were already being applied across various services. Therefore, we needed to understand the major problems that exist in social recommender services that impose designing unique privacy enhancing framework to mitigate or avoid some of them. Good privacy enhancing framework should be able to elaborate in any kind of social recommender services and produce accurate referrals. In this section, the challenges will be presented in designing the proposed collaborative privacy framework.

- Attaining data privacy: The main concern for the users while utilizing an external recommender service is to preserve the privacy of their sensitive data during the recommendation process. Users expect to define what data to release or share for the service and what data to hide. A good privacy enhancing framework should allow the users to specify their own privacy requirements and control what to share over their data. Moreover, the released data should be concealed somehow in order not to be linked to its original version. Additionally, the proposed privacy enhancing framework should attain privacy by combining efforts from both technical and legal domains.
- Accuracy of results: When the users utilize an external recommender service, they expect to receive accurate results. In order to achieve this, a good privacy enhancing framework should seek to diminish variance and bias of the data. Additionally, the good privacy enhancing framework should provide parameters to control the level of concealing of released data, to implicitly inform the user about the expected accuracy he/she might get in return to this desired level.
- Diversity of users' profiles: With the exponential daily growth in the number of items offered by the majority of IPTV content providers, users usually are exposed to a small proportion of items in relation to the total number of items. An effect for that, users' profiles become sparse which can cause difficulties in measuring similarities between users and the execution of any concealment process properly. A good privacy enhancing framework should weigh this problem and try to mitigate its effects on both of the concealment process and recommendations.
- Redundant Items: In numerous IPTV content providers, it might occur that different names might return to the same content. Unreasonable referrals can be generated if this problem exists. Moreover, synonyms can cause a severe privacy invasion, as an individual user might set a rule to prevent the release of a certain item, but this item might have different names within the content provider. A good privacy enhancing framework should consider a way to mitigate the effect of the synonym problem on privacy invasion.
- User Centric Collection: New users of the IPTV content provider cannot join a recommendation process without having a sensible profile. Current social recommender services force the new users to rate a predefined set of items. Although this problem diminishes after a period of system usage, new users will not be able to receive referrals during this period. A good privacy enhancing framework should consider methods to allow new users

to receive referrals based on querying other users in order to reduce the waiting period and complete their profiles quickly.

2.3 Related works

The majority of the existing recommender systems are based on collaborative filtering, while the others focus on content based filtering using EPG data [1]. Collaborative filtering techniques build users' profiles in two ways, either upon ratings (explicit rating procedures) or log archives (implicit rating procedures) [21]. These procedures lead to two different approaches for the collaborative filtering including the rating based approaches and log based approaches. The majority of the literature addresses the problem of privacy on collaborative filtering techniques, due to the potential source of leakage of private information shared by the users as shown in [39]. In [20] It is proposed a theoretical framework to preserve privacy of customers and the commercial interests of merchants. Their system is a hybrid recommender system that uses secure two party protocols and public key infrastructure to achieve the desired goals. In [4, 5] it is proposed a privacy preserving approach based on peer to peer techniques using users' communities, where the community will have an aggregate user profile representing the group as whole but not individual users. Personal information will be encrypted and the communication will be between individual users but not servers. Thus, the recommendations will be generated at a client side. In [45, 46] it is suggested another method for privacy preserving on centralized recommender systems by adding uncertainty to the data using a randomized perturbation technique while attempting to make sure that necessary statistical aggregates such as the mean don't get disturbed much. Hence, the server has no knowledge about true values of individual rating profiles for each user. They demonstrate that this method does not decrease essentially the obtained accuracy of the results. But a recent research work [26, 30] pointed out that these techniques don't provide levels of privacy as it was previously thought. In [30] it is pointed out that arbitrary randomization is not safe because it is easy to breach the privacy protection it offers. They proposed a random matrix based spectral filtering techniques to recover the original data from the concealed data. Their experiments divulged that in various cases random perturbation techniques preserve very little privacy. Similar limitations were detailed in [26]. Storing users' rating profiles on their own side and running the recommender system in distributed manner without relying on any server is another approach proposed in [40], where authors proposed transmitting only similarity measures over the network and keep users rating profiles a secret on their side to preserve privacy. Although this method eliminates the main source of threat against users' privacy, it requires higher cooperation among users to generate useful recommendations. The work in [50] stated that existing similarities deem useless as traditional user profiles are sparse and insufficient. Recommender systems need new ways to calculate user similarities. They have used trustworthiness to define the relationship between two users. The authors in [22] show the correlation between similarity and trust and how it can elevate recommendations accuracy.

2.4 The attack model

In this work, the collaborative privacy framework preserves the privacy of user rating profile from the attack model presented in [44]. The attack model for data concealment techniques is different from the attack model for encryption-based techniques, but no common standard has been implemented for data concealment. Existing attack models have primarily considered a case where the attacker correlates the concealed data with data obtained from other publicly-accessible databases in order to reveal the sensitive information. But the attack model presented in [44], considers a case where the attacker colludes with some users in the network to obtain some partial information about the process used to conceal the data and/or some of the original data items themselves. The attacker can then use this partial information to attempt to reverse engineer the entire dataset. Hence, in this paper, we can define concealment as a technique that enables the user who wants recommendations in a network of users, to conceal his/her raw ratings in his/her profile during the recommendations process, such that, the other users in the network and social recommender service cannot learn any ratings in his/her raw profile. The threat model for the adversaries assumed in this paper is honest-but-curious model. Where the adversary aims to collect users profiles in order to identify and track them. Thus, we consider our main adversary to be an untrusted social recommender service; moreover we do not assume social recommender service to be completely malicious. This is a realistic assumption because the recommender service needs to accomplish some business goals and increase its revenues. The social recommender service can construct the profiles of the users based on the shared interests between various users. Hence, the problem we are tackling is to detain the ability of the adversary to identify items' ratings of the users and to prevent the malicious recommender service from recognizing items in these profiles. Intuitively, the system privacy is high if the malicious users and the recommender service are not able to reconstruct the real profiles of users.

3 A scenario for the collaborative privacy framework in IPTV content distribution network

We extend the scenario proposed in [13–17], where a social recommender systems (PRS) is implemented as an external third-party service and the users of a specific IPTV content provider are giving their rating profiles to this external recommender service in order to receive recommendations. The basic idea for recommendation process within our collaborative privacy framework is as follows: Upon receiving a request from the target user (the user who is requesting recommendations for specific genre), different coalitions of users who are willing to participate in this recommendation are formed, where each set of coalesced users is named "peer-group". Each peer-group is managed by an elected peer, which will be named super-peer, where each super-peer within a peer-group will be acting as a trusted aggregator for the data collected from the members of its peer-group. Additionally, this super-peer will be responsible for anonymously sending the group profile to the social recommender service, where each group profile contains the aggregated data of members within the peer-group of this super-peer. Finally, after receiving the referrals list, the super-peer will be responsible for distributing this list back to the members of its peer-group.

Each user who wishes to participate in the recommendation process conceals his/her ratings' profile using the local concealment process provided by our framework, such that each profile is concealed in a manner to prevent the super-peers from learning the raw ratings of each participant. The super-peer collects these concealed rating profiles from various members in the peer-group and then it computes an aggregation on them in order to create a group profile, which does not expose the individual ratings. Next, the group profile is encapsulated using the global concealment process at the super-peer side, then after, this globally concealed group profile is submitted to the social recommender service to predicate ratings for the recommended items. The referrals list will be offered in the end to the target-user and the members of peer-groups. The collaborative filtering task at the social recommender service (PRS) will perform its predication phase on the globally concealed group profiles without exposing the raw data of each user.

Figure 2 shows the architecture of our approach. Our solution relies on hierarchical topology proposed in [42]; where participants are organized into peer-groups managed by an elected super-peers. Electing super-peers is based on negotiation between group members and security authority centre . Security authority centre (SAC) is a trusted third party that is responsible for generating certificates for the target-user and super-peers, and managing these certificates. In addition, SAC is responsible for making assessment on those super-peers according to users' reports and periodically updates the reputation of these super-peers. The reputation mechanisms are employed to elect suitable super-peers based on estimating values for user-satisfaction, trust level, processing capabilities and available bandwidth, detailed and complex reputation mechanisms can be found in [6]. When a problem with specific super-peer occurs during the recommendation process, an end-user can report it to SAC. After investigation, the assessment of the super-peer will be degraded. This will limit the chance for electing it as a super-peer in the future. On the other hand, successful recommendations processes will help to upgrade the super-peer reputation. IPTV content providers can offer certain benefits for those users who have sustainable success rate as super-peers (like free content, prizes, discount coupon … etc.). So, in order to summarize, when the target user broadcasts a message to other users in the IPTV content distribution network to request recommendations for a specific genre or category of items. Peer-groups are formed with specific users who are interested to get recommendations in regard to this request. The peer in the peer-group with the highest reputation is elected as a "super-peer".

Our solution depends upon the set top-box (STB) device at the user side. The STB is an electronic appliance that connects to both the network and the home television. With the advancement of data storage technology each STB is equipped with a mass storage, e.g. Cisco STB. The proposed EMCP components are hosted on that STB. On the one hand, the STB storage stores the user rating profile. On the other hand, the social recommender service (PRS) maintains a centralized rating database that is used to provide recommendations to the target user if the number of responding users is below a certain threshold. Moreover, this centralized rating database is sufficient enough for building and training the recommendation model and supporting the IPTV content distribution providers from different perspectives, such as maximizing the precision of target marketing and improving the overall performance of the current distribution network by building up an overlay to increase content availability, prioritization and distribution based on the predicated recommendations. We alleviate the user's identity problems by using anonymous pseudonyms identities, which are obtained from SAC.

4 The proposed enhanced middleware for collaborative privacy (EMCP)

In the beginning, we want to introduce the notion of privacy within our framework; we need to justify what we mean by privacy first. Privacy is an elusive concept that is difficult to be outlined, it is not an entirely technical subject but it is connected to aspects of legislation, service providers' policies, and social norms. Privacy is an adjustable notion depending on the users' perception of risk and profit. Some users could reveal their personal information if they are given advantages in return. These advantages can be in the form of a discount coupon, accurate referrals, and personalized content. However, when something is considered private to the user, it usually means there is something within them that is considered inherently personally sensitive to avoid discrimination, personal embarrassment, or harm to their professional reputations. The degree to which private information is exposed thus depends on how the public will receive this information, which differs between situations and over time. The research in [38] conceptualizes privacy as the "selective control of access to the self" regulated as dialectic and dynamic processes that embrace multi-mechanistic optimizing behaviours. Moreover, The work presented in [49], in which privacy

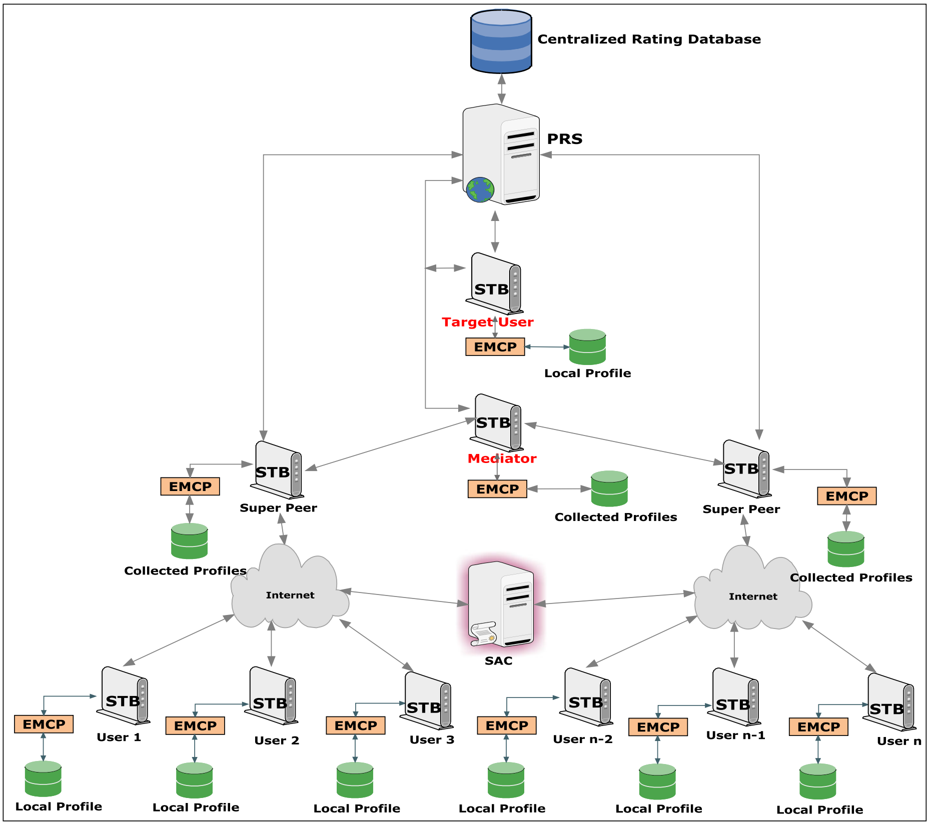


Fig. 2 IPTV content distribution network with our collaborative privacy framework

was defined as permitting the external services to extract a valid knowledge without learning the underlying users' personal data. We can define the notion of privacy as follows: "A target user in a network of users wants recommendations about specific items. However, the target user does not have to reveal the real ratings in his/her profile during the recommendations process and other users in the network cannot learn any real ratings in his/her profile". Thus, the privacy view within the collaborative privacy framework is surrounding the disclosure of users' rating profiles which can be considered as the backbone of our solution. In the next sections, we will present the proposed *EMCP* middleware for protecting the privacy of users' rating profiles, where this middleware serve as the cornerstone in our collaborative privacy framework.

Figure 3 illustrates the components of the proposed enhanced middleware for collaborative privacy (*EMCP*) running inside the user's STB. *EMCP* consists of different co-operative agents. A learning agent captures the user's interests about miscellaneous items explicitly or implicitly to build a rating database and meta-data database. The local obfuscation agent implements a local concealment process to achieve user privacy while sharing his/her preferences with super-peers or the external social recommender service (PRS). The encryption agent is only invoked if the user is acting as a super-peer in the recommendation process; it executes global concealment on the group profile (collected profiles from the members of the peer group). The two stage concealment process acts as wrappers to conceal preferences before they are shared with any external social recommender service.

Since the database is dynamic in nature, the local obfuscation agent periodically conceals the updated preferences, and then a synchronize agent forwards them to the social recommender service (PRS) upon owner permission. Thus, recommendation can be made on the most recent preferences. Moreover, the synchronize agent is responsible for calculating and storing parameterized paths in an anonymous network that attain high throughput, which in turn can be used in submitting preferences anonymously. The policy agent is an entity in *EMCP* that has the ability to encode privacy preferences and privacy policies as XML statements depending on the host role in the recommendation process. Hence, if the host role is as a “super-peer”, the policy agent will have the responsibility to encode the data collection and data usage practices as P3P policies via XML statements which are answering questions concerning the purpose of collection, the recipients of these profiles, and the retention policy. On the other hand, if the host role is as a “participant”, the policy agent acquires the user’s privacy preferences and expresses them using APPEL as a set of preferences rules which are then decoded into a set of elements that are stored in a database called “privacy preferences” in the form of tables called “privacy meta-data”. These rules contain both a privacy policy and an action to be taken for such a privacy policy, in such a way this will enable the preference checker to make self-acting decisions on objects that are encountered during the data collection process regarding different P3P policies (e.g.- privacy preferences could include: certain categories of items should be excluded from data before submission, expiration of purchase history, usage of items that have been purchased with the business credit card and not with the private one, generalize certain terms or names in the user’s preferences according to a defined taxonomy, using synonyms for certain terms or names in the user’s preferences, suppressing certain items from the extracted preferences, and inserting dummy items that have the same feature vector like the suppressed ones as described in [18], limiting the potential output patterns from extracted preferences etc. in order to prevent the disclosure of sensitive preferences in the user’s profile). Query rewriter rewrites the received request constrained by the privacy preference for its host. An overview of the recommendation process in the proposed framework operates as follows:

1. The learning agent collects the user’s ratings regarding different items which have exposed to him/her; the output of this step represents the user’s local profile. The local profile is stored in two databases, the first is the rating dataset which contains (item_id,

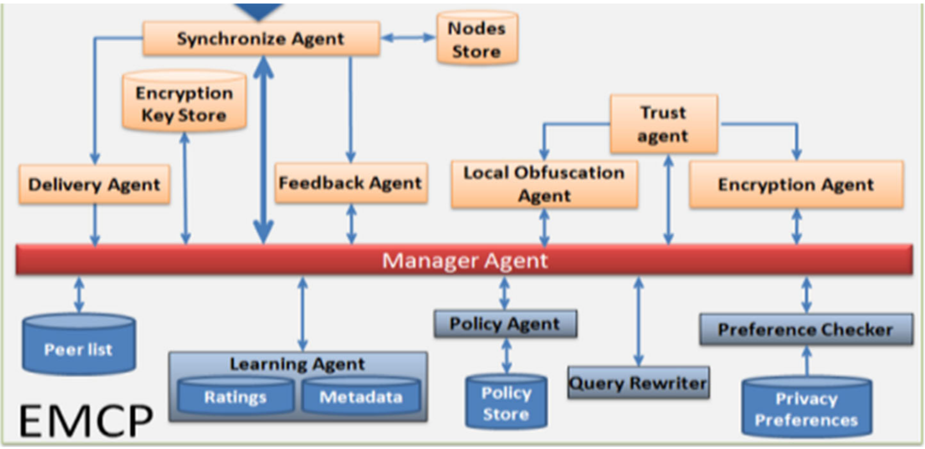


Fig. 3 *EMCP* components

rating) and the second is the meta-data dataset which contains feature vector for each item such as [18] (item_id, feature1, feature2, feature3). The feature vector can include genres, directors, actors and so on. Both implicit and explicit ways for information collection [32] are used to construct these two databases and maintain them.

2. As stated in [14], the target user broadcasts a message to other users in the IPTV content distribution network to request recommendations for specific genre or category of items. Then he/she invokes the local obfuscation agent to execute a local concealment process in order to conceal a set of items' ratings in his/her local profile that is related to this genre or category. In order to hide the items identifiers and meta-data from other users, The manger agent uses locality-sensitive hashing (LSH) [27] to hash these values. One interesting property for LSH is that similar items will be hashed to the same value with a high probability. Super-peers and PRS are still be able to perform computation on the hashed items using appropriate distance metrics like hamming distance or dice coefficient. Finally, the target user dispatches these concealed items' ratings along with their associated hashed values to the Individual users who decided to participate in this recommendation process.
3. Each group of users negotiates with the external SAC to select a peer with the highest reputation between them to act as a "super-peer" which will act as a communication gateway between the target user, PRS and users in its underlying peer-group.
4. Each super-peer negotiates with both the target user and the recommender service to express its privacy policies for the data collection and usage process via P3P policies.
5. At each participant side, the manager agent receives the request from the target user along with the P3P policy from the elected super-peer; then it forwards this P3P policy to the preference checker and the request to the query rewriter. The preference checker ensures that the extracted preferences do not violate the privacy of its host which were previously decaled by the use of APPEL preferences. The query rewriter rewrites the received request based on the feedback of the preference checker. The modified request is directed to the learning agent to start the collection of preferences that could satisfy the modified query and forwards it to the local obfuscation agent. Finally, the policy agent audits the original and modified requests and P3P policy with previous requests in order to prevent multiple requests that might extract sensitive preferences.
6. The local obfuscation agent at each participant side executes the local concealment process on the preferences extracted by the learning agent from step 5 and the manager agent hashes their identifiers and meta-data using LSH. Thereafter, these locally concealed data from members in the peer-group are submitted to the super-peer of this group, to be aggregated in a group profile, which does not expose the raw individual ratings for the members. Secure routing protocols can be utilized to hide the network identities of group members when submitting their locally obfuscated profiles to the super-peers.
7. Each super-peer builds a group profile by collecting users' pseudonyms and aggregating items' ratings such that all the <hashed value, rating> elements belonging to similar items clustered together. This allows computing items' popularity curve at each super-peer. After the construction of the group profile, the super-peer invokes the encryption agent to execute the global concealment process on the group profile of its peer-group. Then after, the globally concealed group profile is submitted to the PRS in order to predicate ratings for the recommended items in the referrals list. The super-peer can seamlessly interact with the social recommender service (PRS) by posing as a user and has the globally concealed group profile as his/her own profile.
8. Upon receiving the globally concealed group profiles from all super-peers, the social recommender service (PRS) stores these profiles along with their users' pseudonyms and hashed values in the centralized rating database. After that, PRS executes the collaborative

filtering algorithm on the received globally concealed group profile in order to extract a referrals list. Then after, PRS forwards the generated referrals list along with the predicated ratings to each super-peer in the peer-group. Each super-peer publishes the final list to the target user and/or the members of its peer-groups. Finally, each member reports scores about the elected super-peer of his/her peer-group and the target-user to SAC, which helps to determine the reputation of each entity involved in the referrals generation.

4.1 The concealment process in the collaborative privacy framework

In the next sections, we will present the two stage concealment process used in *EMCP* to conceal the user rating profile in a way that secures users' ratings in an untrusted PRS with minimum loss of accuracy. In our framework, each user has two datasets representing his/her profile. Local profile which represents the actual raw ratings' profile of the user for different items; it is stored on his/her STB. When the end-user participates in a recommendation process, he/she locally conceal this local profile before sending it to the super-peer. While the public profile represents the output of the two-stage concealment process where the user gets the required recommendations directly from the PRS based on this profile. We had performed experiments on real datasets to illustrate the applicability of our two stage concealment process along with the privacy and accuracy levels achieved using this process.

A closer look to the attack model proposed in [41] reveals that, if the released set of users' preferences are fully distinguishable with respect to some features from other users' preferences in the group profile, this user can be identified if an attacker manages to correlate the revealed preferences with preferences obtained from other publicly-accessible databases. Therefore, it is highly desirable that at least a minimum number of items in the released preferences to have a similar features' vector close to each real item preferred by the user. A real item can be described by a features' vector, which includes genres, directors, actors and so on. Both implicit and explicit ways for information collection can be used to construct this features' vector. Moreover, the data sparsity problem associated with the group profile can be used to formulate some attacks as shown in [41]. The main aim for the global concealment process is to alleviate the data sparsity problem by filling the unrated cells in the aggregated preferences in such a way to improve recommendation accuracy and increase the attained privacy. Before going into details of the two stage concealment process, we will introduce a couple of relevant definitions.

Definition 1 (Dissimilarity measure) This metric measures the amount of divergence between two items with respect to their features' vectors. We use the notation $\mathcal{D}_m(I_u, I_n)$ to denote the dissimilarity measure between items I_u and I_n based on the features vector of each one of them.

$\mathcal{D}_m(I_u, I_n) < \delta \Rightarrow I_u \tilde{} I_n$ [I_u is similar to I_n], δ is a user defined threshold value.

Definition 2 (Affinity group) The set of items that are similar to item I_u with respect to *pth* attribute A_p of features vector is called affinity group of I_u and denoted by $C_{A_p}(I_u)$.

$$\begin{aligned} C_{A_p}(I_u) &= \left\{ I_n \in D_n \middle| (I_u \tilde{} I_n) \wedge (A = A_p) \right\} \\ &= \left\{ I_n \in D_n \middle| \mathcal{D}_m(I_u, I_n) < \delta \right\} \end{aligned}$$

Definition 3 (K-Similar item group) Let D_ϖ be the real items dataset and $\widetilde{D_\varpi}$ is its locally concealed version. We say $\widetilde{D_\varpi}$ satisfies the property of k-Similar item group (where K value is defined by the user) provided for every item $I_u \in D_{\varpi}$. There is at least *k-1* other distinct fake items $I_{n_1}, \ldots I_{n_{(k-1)}} \in D_n$ forming affinity group such as:

$$FV\ (I_{n_i})\ \widetilde{}\ FV\ (I_u),\ \forall\ 1 \leq i \leq k-1$$

4.1.1 Local concealment using clustering based obfuscation (CBO) algorithm

We propose a novel algorithm for concealing the user's preferences before sending them to the super-peer. This algorithm is called clustering based obfuscation (CBO), which has been designed especially for the sparse data problem existing within the user's profile. Our motivation to propose the CBO algorithm is the limitation of the current obfuscation models that fail to provide an overall anonymity as they don't consider matching items based on their features' vectors. CBO uses the feature vectors of the existing real items to select a set of fake items similar to these real items in order to create homogeneous concealed dataset. Using fake transactions to maintain privacy was presented in [36, 37], the authors considered adding fake transactions to anonymise the original data transactions. This approach has several advantages over other schemes including that any off-the-shelf recommendation algorithms can be used for analysing the locally concealed data and the ability of providing a high theoretical privacy guarantee. The locally concealed dataset obtained using CBO should be indistinguishable from the original dataset in order to preserve privacy. Moreover, auditing sub-queries for every recommendation request allows the CBO to protect the sequential release of the real items by controlling the amount of fake items to be added to the released ones. The core idea for CBO is to split the dataset into two subsets, the first subset is modified to satisfy the K-Similar item group definition, and the other subset is concealed by substituting the real items with fake ones based on a probabilistic approach. The CBO algorithm creates a concealed dataset D_P based on the following steps:

1. Suppress the set of sensitive items from the local profile based on the user's preferences. Thereafter we will have a dataset D as the real dataset.
2. Selecting a percent ϖ of highly rated items in the dataset D to form a new subset D_{ϖ}. This step aims to reduce the substituted fake items inside the concealed dataset D_P. Moreover, it maintains data quality by preserving the aggregates of highly rated items.
3. CBO builds affinity groups for each real item $\forall\ I_u \in D_{\varpi}$ through adding fake items to form K-Similar items group. We implemented this task as a text categorization problem based on the features vectors of real items. We have used the bag-of-words naive Bayesian text classifier [35] that extended to handle a vector of bags of words; where each bag-of-words corresponds to a feature (title, cast, genre, etc.). We have trained the classifier to learn a model from a set of real items. Then we used this model to select a set of fake items and predict their ratings. The task continues until every real item in D_{ϖ} have various affinity groups associated with it, and then in the end we get the new dataset $\widetilde{D_{\varpi}}$.
4. For each $I_u \in D_u = D - D_{\varpi}$, CBO selects a real item $\{I_u\}$ from a real item setD_u with probabilityα or selects a fake item $\{I_n\}$ from the candidate fake item set D_n with probability $1-\alpha$. The selected item I_P is added as a record to the concealed dataset D_P. This method achieves the desired privacy guarantee because the type of selected item and α are both unknown to the super-peer and PRS. Each user can decide locally the values of these parameters. The process continues until all real items in D_u are selected. Each item $I_P \in D_P$ follows: $P(I_P) = \alpha P(I_u) + (1-\alpha) P(I_n)$
5. Finally, the concealed dataset D_P is merged with the subset $\widetilde{D_{\varpi}}$ that was obtained from step 3.

4.1.2 Global concealment using random ratings generation (RRG) algorithm

After ending the execution of the CBO algorithm, each user submits his/her locally concealed data to the super-peer of his/her peer-group. The super-peer aggregates the data collected from these members, then after, it starts building a group profile. The super-peer invokes the encryption agent to globally conceal the group profile using random ratings generation (RRG) algorithm. The super-peer will not be able to know the real items in the aggregated datasets as these items already concealed using CBO algorithm. The main aim for the RRG is to alleviate data sparsity problem by filling the unrated cells in such a way to improve recommendation accuracy at PRS side and increase the attained privacy for the users. The RRG algorithm consists of following steps:

1. The encryption agent finds a number of majority rated items I_r and partially rated items by all users $I-I_r$, where I denotes the total number of items in the aggregated datasets.
2. The encryption agent randomly selects an integer ρ between 0 and 100, and then chooses a uniform random number ξ over the range $[0,\rho]$
3. The encryption agent decides a percent ξ of the partially unrated items in the aggregated datasets and uses the *KNN* algorithm to predicate the values of the unrated cells for that percentage as in [24]. Then, the remaining unrated cells are filled by random values chosen using a distribution reflecting the ratings in the aggregated datasets (average and variance). The encryption agent should select ρ in a way to achieve the required balance between privacy and accuracy.

4.2 The main characteristics of the collaborative privacy framework

The current privacy legislations rely on the commitment of the IPTV content providers on revealing their data handling practices accurately. However, the current perspective illustrates that it is likely for them to not follow these practices in full. The proposed framework reduces privacy risks and facilitates privacy commitment. Moreover, it realizes privacy aware recommendations while complying with the current business model of third-party social recommender service. The main characteristics obtained through the proposed collaborative privacy approach are as follows:

1. **Collection Method**: The proposed solution attains an explicit data collection mode. Users are aware that data collection within a recommendation process is happening and they can make a wise decision about whether or not to provide their data in this recommendation process. Privacy policies such as P3P are utilized to explain to the users how their data is going to be used. Users utilize privacy preferences in order to control what data from their profiles gets collected in which concealment level. However, formalizing such privacy preferences is not an easy task. Users need to realize various privacy issues. Additionally, users need to deduce future recommendation requests that might raise privacy concerns for their collected data. The user can employ an anonymous network while sending this locally concealed data to either the super-peer or the social recommender service.
2. **Duration**: The proposed solution attains a session based collection that allows for a simpler service that does not need the storage and retrieval of users' profiles. The data related to the recommendation process is collected from users' profiles in a concealed form. This concealed data is only feasible for recommendation purposes. This reduces the privacy concerns as minimal data to be collected and also ensures the compliance with the

privacy laws. The concealed data is stored at the third party service in order to enhance the recommendation model and future requests. Moreover, this data by default is protected by the privacy protection laws.

3. **Initiation**: The proposed solution attains a user based recommendation. Users are the entities that initiate the recommendation process; each user in the network is aware that a recommendation process is happening and he/she can decide whether or not to join it. The incentive for participants when joining a recommendation request includes receiving referrals regarding a certain topic in a private manner.
4. **Anonymity**: The proposed solution attains anonymity which aids in preventing frauds and sybil attacks. The anonymity is realized within the collaborative privacy framework using the following procedures:

   a. Dividing system users into a coalition of peer-groups: each peer-group to be treated as one entity by aggregating its members' concealed data in one aggregated profile at the super-peer, then this super-peer will handle the interaction with the social recommender service. Participants within the coalition interact with each other in a P2P fashion and form a virtual topology to aggregate their data.
   b. Using anonymous channels like Tor: Individual participants might benefit from these anonymous channels while contacting the recommender service or other members in their coalition.
   c. Utilizing pseudonyms for users: each user within the system is identified by a pseudonym in order to reduce the probability of linking his/her collected profiles' data with a real identity.
5. **Local Profiles**: Our solution attains local profiles storage. Users' profiles are stored locally on their own devices (Setup box, Smart phone, Laptop…) in an encrypted form. This can guarantee that these profiles are attainable only to their owners. Furthermore, in doing so these profiles will be inaccessible to viruses or malware that may affect the user's machine to gather his/her personal data. As a result, each user will possess two profiles; one is a local profile in a plain form that is stored locally in his/her machine and it is updated frequently. The other is a public profile in a concealed form that is stored remotely at the service provider and it is updated periodically within each recommendation process where this user participated.
6. **Two Stage Concealment Process**: Our solution relies on a concealment process which is carried out in two consecutive steps in order to make sure that the data does not leave the personal device of the end-user until it is properly concealed. The proposed techniques destroy the structure in data but, at the same time, maintain some properties in it which are required in the planned recommendation process. Additionally, the implementation of such applications confirmed that is feasible to make use of and, at the same time, to protect the personal sensitive data of individuals, and do so in an accurate way.

5 Proof of security for the two stage concealment process

5.1 Proof of security for CBO algorithm

Differential privacy is a new privacy notion proposed in [11] which provides a strong privacy guarantees that is independent of the auxiliary information that an attacker might have. It assumes that any results of private database should not significantly change with the addition, or update of a single record. In this paper, the input of CBO is the raw ratings of a user who participates in the recommendation process, and the output will be the locally concealed profile, which will be delivered to either the super-peer or the PRS.

Definition 4 [11] (Differential Privacy) A privacy mechanism $\mathscr{M}:\mathscr{R}^n \rightarrow \mathscr{R}^n$satisfies -differential privacy if for all data sets $D_1, D_2 \in \mathscr{R}^n$differing in at most one element, and for all possible $\widetilde{D} \in Range(\mathscr{M})$.

$$Pr\left[\mathscr{M}(D_1) = \widetilde{D}\right] \leq e * Pr\left[\mathscr{M}(D_2) = \widetilde{D}\right]$$

This probability is taken over the randomness of the mechanism M. e is a mathematical constant that represents the base of the natural logarithm. The smaller values of e correspond to higher privacy levels. Several work in the literature have been done to construct differentially private mechanisms through perturbations using laplacican mechanism [3, 43].

Definition 5 [2] (Differential Privacy in Local Model) A privacy mechanism $\mathscr{M}(H) = (v(h_1), \ldots, v(h_n))$ such that $v: R \rightarrow R$ and $H = (h_1, h_2, \ldots. h_n)$ satisfies -differential privacy if for all items $I, I' \in R$ and for all possible $o \in Range(R)$

$$Pr[v(I) = o] \leq e * Pr\left[v\left(I'\right) = o\right]$$

This probability is taken over the randomness of the views v. This definition is applied if $\mathscr{M}(H)$ can be decomposed to a set of views $(v(h_1), \ldots, v(h_n))$, where each view has an independent private variable $(h_1, h_2, \ldots. h_n)$.

In the context of CBO algorithm, with probability α, CBO selects a real item $\{I_u\}$ from real item set D_u or with probability $1-\alpha$, CBO selects a fake item $\{I_n\}$ from the candidate fake item set D_n. The selected item I_P is added as a record to the released dataset D_P. This is known as the randomized response method [12]. As a result, we re-apply the above equation in CBO as follows:

$$Pr[\mathscr{M}(\mathrm{I_u}) = \mathrm{I_P}] \leq e * Pr[\mathscr{M}(\mathrm{I_n}) = \mathrm{I_P}]$$

To achieve - differential privacy for CBO, we need to find an optimal probability for α as a function of , where the smaller values for α, the more accuracy preserved in the released dataset. Thus, the value of α must be in:

$$\frac{1}{(e + 1)} \leq \alpha \leq \frac{1}{2}$$

Due to that, CBO bounds the information the adversary gets when it receives an item within the locally concealed profile. The adversary knows that the end-users released a real item with probability α.

Theorem: (Privacy guarantees for an item in D_p) Setting the probability α to $\frac{1}{(e+1)}$ satisfies Definition 2 and guarantees -differential privacy for the datasets obfuscated with CBO.

We divide real item set D_u into two parts, D_u^s is a binary set of selected items' positions in D_p^s and D_u^{-s} is a binary set of items' positions that were not selected in D_p^{-s}. A similar partitioning to the item set D_n^s and D_n^{-s}. Let $\beta = 1 - \alpha$

$$\begin{aligned}\left|\ln\frac{Pr[\mathscr{M}(\mathrm{D_u}) = \mathrm{D_P}]}{Pr[\mathscr{M}(\mathrm{D_n}) = \mathrm{D_P}]}\right| &= \left|\ln\frac{Pr\left[\mathscr{M}\left(\mathrm{D_u^s}\right) = \mathrm{D_p^s}\right]Pr\left[\mathscr{M}\left(\mathrm{D_u^{-s}}\right) = \mathrm{D_p^{-s}}\right]}{Pr\left[\mathscr{M}\left(\mathrm{D_n^s}\right) = \mathrm{D_p^s}\right]Pr\left[\mathscr{M}\left(\mathrm{D_n^{-s}}\right) = \mathrm{D_p^{-s}}\right]}\right| \\ &= \left|\ln\frac{Pr\left[\mathscr{M}\left(\mathrm{D_u^s}\right) = \mathrm{D_p^s}\right]}{Pr\left[\mathscr{M}\left(\mathrm{D_n^s}\right) = \mathrm{D_p^s}\right]}\right| \text{Where } \mathscr{M}\left(\mathrm{D_u^{-s}}\right) = \mathscr{M}\left(\mathrm{D_n^s}\right) \\ &= \left|\ln\frac{\beta^i\alpha^v}{\beta^v\omega}\right|\end{aligned}$$

Where i is the number of ones in the dataset and v is the number of zeros in the dataset. The value $\alpha^i \leq \omega \leq \beta^i$ and $0 \leq \upsilon \leq i$. So if $\alpha=1/(e^{\in/i}+1)$ then $\beta=(1-\alpha)=(e^{\in/i})\alpha$ and $\alpha/\beta=e^{-\in/i}$. Thus, $\alpha^i \leq \omega \leq (e^{\in})\alpha^i$ or in other words $1 \leq \omega/\alpha^i \leq (e^{\in})$ which implies that $1 \geq \alpha^i/\omega \geq 1/(e^{\in})$ and then $0 \geq \ln\alpha^i/\omega \geq e^{\in}$

$$\left|\ln\frac{\beta^i\alpha^v}{\beta^v\omega}\right| = \left|\ln\frac{(e\ )\alpha^i}{\omega}\frac{1}{e^{v/i}}\right| = \left|\ln\left(\alpha^i/\omega\right) + e(i-v)/i\right|$$

The term $0 \leq e(i-v)/i \leq \in$, thereafter the maximum value of $|\ln(\alpha^i/\omega)+e(i-v)/i|$ is $\in$. Additionally, in our case, CBO ensures the privacy for each item in the real item set D_u, this we need to compute the values of α that ensure differential privacy for each item. Hence, we will set $i=1$, so it will be easy to verify that $\alpha\frac{1}{(e^{\in}+1)}$ is the minimum value that attains $\left|\ln\frac{pr[M(I_u)=Ip]}{pr[M(I_n)=Ip]}\right| \leq \in$. Consequently, this proves our theorem.

5.2 Proof of security for RRG algorithm

The social recommender service cannot figure out the real rated items within the group profile due to the random filling of the unrated cells with a set of random values extracted from a distribution reflecting the ratings in the aggregated datasets. The PRS might try to infer the randomly rated cells. However, the probability of correctly estimating ρ is 1 out of 100, as ρ is an integer between 0 and 100. After obtaining ρ, the probability of correctly estimating ξ is 1 out of ρ, as ξ is a uniform random number selected over the range $[0, \rho]$. After filtering out the randomly filled unrated cells and then determining the items which are corresponding to them. The attacker can determine the existing real items γ_{rel} which have been collected from the members of the group. If we assume that the rating scale for each item is from 1 to 4. The attacker knows that this item might be belonging to the first quarter $\frac{\gamma^1_{rel}}{4}$ of items which has the value 1 or the second quarter $\frac{\gamma^2_{rel}}{4}$ of items which has the value 2 or the third quarter $\frac{\gamma^3_{rel}}{4}$ of items which has the value 3, or the final quarter $\frac{\gamma^4_{rel}}{4}$ of items which has the value 4. The probabilities of correctly predicting which items have been rated either 1, 2, 3, or 4 are 1 out of $C^{\gamma^1_{rel}}_{\gamma^1_{rand}}$, $C^{\gamma^2_{rel}}_{\gamma^2_{rand}}$, $C^{\gamma^3_{rel}}_{\gamma^3_{rand}}$ and $C^{\gamma^4_{rel}}_{\gamma^4_{rand}}$ where $\gamma^1_{rel}, \gamma^2_{rel}, \gamma^3_{rel}$, and γ^4_{rel} represent the items with ratings 1, 2, 3, and 4. Respectively, $\gamma^1_{rand}, \gamma^2_{rand}, \gamma^3_{rand}$, and γ^4_{rand} show the items with randomly filled ratings of 1, 2, 3, and 4. Thus, the probability of correctly filtering out an unrated item is 1 out of $\left[100 \times \xi \times C^{\gamma^1_{rel}}_{\gamma^1_{rand}} \times C^{\gamma^2_{rel}}_{\gamma^2_{rand}} \times C^{\gamma^3_{rel}}_{\gamma^3_{rand}} \times C^{\gamma^4_{rel}}_{\gamma^4_{rand}}\right]$. Then after, PRS will be able to filter out the existing collected items. However, if the PRS decided to collaborate with a malicious peer in order to reveal the raw ratings of the victim user, the privacy guarantees that CBO was designed to be a robust shield against this attack. As a result, the PRS will not be able to learn the raw ratings of the victim user. Moreover, the users' raw ratings are locally concealed independently in various peer-groups. Thus, with the growing number of peer-groups, the attained privacy level is increased.

6 Experiments

In this section, we will describe the implementation of our proposed collaborative privacy framework. The experiments run in Intel® Core 2 Duo™ 2.4 GHz processor with 2 GB Ram. We have used MySQL database as data storage. The proposed techniques for our two stage concealment process have been coded in C++. We used message passing interface (MPI) for the distributed memory implementation of *RRG* protocol to mimic a distributed reliable network of peers. We have evaluated the effect of our proposed two stage concealment process on the concealed data which is used for the recommendations. For that purpose, we have measured that effect regarding two aspects: privacy breach and accuracy of results. The experiments presented here were conducted using the movielens dataset provided by grouplens [34]. The data in our experiments consists of ratings for 36 or more items by 23.500 users. In our dataset, the first column of every raw stores how many items are rated by the user, which is necessary for concealment process. We fixed the number of super-peers to be 3, as described earlier they will be responsible on aggregating the data of 23.496 participants. The recommendation process can be initiated by any user that will be acting as a target-user who is asking for a referrals list. The two stage concealment process between participants is executed locally on their STB devices.

We used the Mean Average Error (MAE) metric proposed in [23] to evaluate the accuracy of the generated predications. MAE is one of the most famous metrics for recommendation quality. We can define it as follows: Given a user predicated ratings set $p=\{p_1,p_2,p_3\cdots p_N\}$ and the corresponding real ratings set $r=\{r_1,r_2,r_3\cdots r_N\}$ MAE is:

$$MAE = \sum\nolimits_{i=1}^{N} {}^{p_i - r_i}\!/_{N}$$

MAE measures the predication variety between the predicated ratings and the real ratings, so smaller MAE means better recommendations provided by PRS. The experiments involve dividing the dataset into a training set and testing set. The training set is concealed then used as a database for PRS. Each rating record in the testing set is divided into rated items t_i $t_{u,k}$ and unrated items $r_i r_{u,i}$. The set t is presented to the PRS for making predication $p_i p_{u,i}$ for the unrated items r_i.

In the first experiment, we wanted to measure the relation between the quantity of fake items in the subset $D\varpi$ (which based on ϖ value and the accuracy of recommendations. We selected a set of real items from movielens, then we split this set into two subsets $D\varpi$ and D_u. We concealed D_u as described before with a fixed value for α to obtain D_p. Then after, we append $D\varpi$ with either items from an optimal fake set or a uniform fake set. Thereafter, we gradually increase the percentage of real items in $D\varpi$ that are selected from movielens dataset from 0.1 to 0.9. For each possible concealment rate (ϖ value), we measured MAE for the whole obfuscated dataset D_p. Figure 4 shows MAE values as a function of the concealment rate. The user selects a concealment rate based on the desired accuracy level required for the recommendation process. We can deduce that with a higher value for the concealment rate a higher accurate recommendation the user can attain. Adding items from the optimal fake set has a minor impact on MAE of the generated recommendations without having to select a higher value for the concealment rate

However, as we can see from the graph, MAE rate slightly decreases in a roughly linear manner with higher values of the concealment rate. Especially, the change in MAE is minor in the range 40 to 60 % that confirms our assumption that accurate recommendations can be provided with less values for the concealment rate. The optimal fake items are so similar to the real items in the dataset, so the local concealment does not significantly change the aggregates in the real dataset and it has a small impact over MAE.

In the second experiment, we seek to measure the impact of adding fake items on the predications accuracy of the various types of ratings. We partitioned the movielens dataset into 5 rating groups. For each rating group, a set of 1300 ratings was separated. The local concealment was applied using optimal and uniform fake sets then the ratings were pre-processed using the global concealment. The resulting datasets were submitted to PRS to perform predications for different rating groups. We repeated the rating predication with different values for α, ϖ, ρ and ξ. Then after, we computed MAE for these predictions. Figure 5 shows the MAE values for the generated predications for each rating group. We can clearly see the impact of adding fake items on the predications of various types of ratings is different. For the optimal fake set, the impact is minor as MAE roughly remains unchanged regardless of the values of α, ϖ, ρ and ξ.

Due to the different levels of privacy concerns between the super-peers, they might select various values for ρ that affect the accuracy for the overall recommendations. This probably influences their revenues, as the IPTV content provider pays for the usage of their content delivery databases, which require a certain quality within the extracted recommendations. To evaluate how the selection of ρ affects the accuracy of recommendations, we performed this third experiment using movielens dataset. We varied ρ from 0 to 100 to show how different values of ξ can affect the accuracy of the predications. Note that when ρ is 0, this means select all unrated items and fill them with random values chosen using a distribution reflecting the ratings in the aggregated datasets. Once we select ρ we can randomly select ξ over the range [0, ρ], after calculating the values of MAE, the results were shown in Fig. 6. As seen from Fig. 6, the accuracy becomes better with augmented ρ values, as the size of the selected portion that was filled using KNN increased and the size of randomized portion decreased. Although, augmenting ρ values attains lower MAE values but we still have a decent accuracy level for recommendations. Accuracy losses result from error in predications such that the predicated ratings might not represent the true ratings for these unrated items. Also there is an error yield

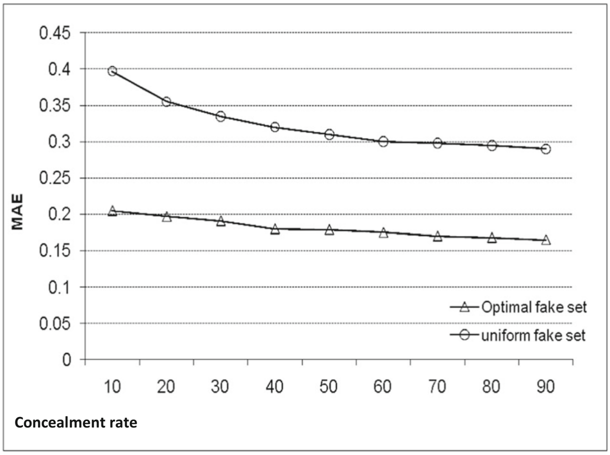


Fig. 4 MAE of the generated predications vs. concealment rate

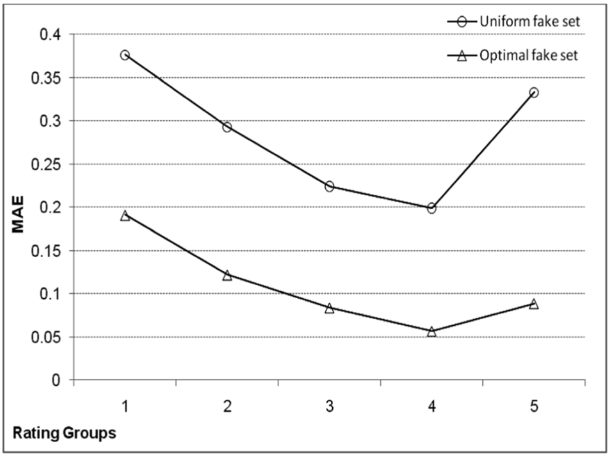


Fig. 5 MAE of the generated predications for ratings groups

from using KNN predications with different values for K. Using these errors; we guarantee a lower limit for privacy breach for the aggregated datasets as shown in Fig. 6. We can conclude that the accuracy losses due to privacy concerns are small and our proposed RRG algorithm makes it possible to offer accurate recommendations.

To mimic attacks on our two stage concealment process, we modelled a PRS trying to link the groups profiles' data to certain users. The user profiles contain a set of items' ratings γ Upon receiving a request for a recommendation process, each user within the peer-group performs local concealment on his/her items' ratings then forward them to the super-peer, who will aggregate all these ratings profiles of the member within the peer-group in one group profile to form γ_{ag}. The super-peer will execute a global concealment on this group profile, then after will forward it to PRS. Each user has a hidden set of items' ratings γ_{hid} and a released set of items' ratings γ_{rel} . The γ_{rel} is already in the group profile at the super-peer side. The total items' ratings by all users can be represented as a bipartite graph, with each user *P* within a peer-group *n* represented as set of nodes

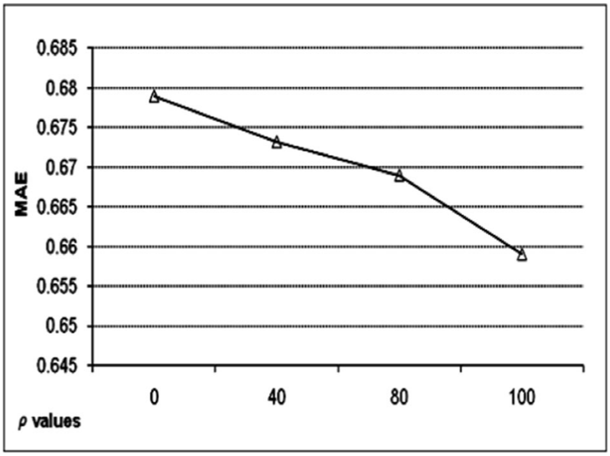


Fig. 6 MAE of the generated predications for different ρ values

P_n and the complete items' ratings set γ. The set of edges connecting a user in P_n to a subset of γ defines the user's profile. The hidden graph, G_{hid}, contains the hidden items' ratings of the user while the release graph G_{rel}, is the graph built by untrusted social recommender service provider colluding with one or more of the super-peers/members. We employ privacy metric proposed in [48] to evaluate the attained privacy by our concealment process. The metric measures the achieved privacy in the two stage concealment process using concepts of graph matching, where PRS tries to match $\mathrm{G_{hid}}$ to $\mathrm{G_{rel}}$. The value S_i represents the frequency of releasing this items' rating for different requests regarding this item's genre. The higher value for this metric implies a higher privacy level attained.

$$privacy = \frac{1}{N} \times \sum_{P_n} \frac{\sum_{i\in\left(\gamma_{rel}/\gamma\right)} \frac{1}{S_i}}{\sum_{i\in\left(\gamma_{rel}\right)} \frac{1}{S_i}}$$

In the fourth experiment, we wanted to show the effect of increasing the number of recommendation requests on the number of users within the peer-group. We extracted the recommendation requests from our training dataset and then we started to forward them to a different user, in order to permit the users to form peer-groups. We selected peer-group No.14 from the created peer-groups to study the effect of varying the number of requests in the group size. In general, decreasing the number of requests decreases the number of users within the peer-group. Figure 7 shows the users involvement in a certain peer-group while varying the number of recommendation requests. As we can conclude from Fig. 7, the average number of users in a peer-group decreases as the number of requests decreases, this is due to the detachment of some of the members from this inactive peer-group and their participation in a request with a different active peer-group. In this case, the users can perform local concealment directly on their items' ratings and send them to the PRS in order to receive recommendations on his/her own. The accuracy of predications will be dependent on the parameters of the local concealment process.

In the fifth experiment, we wanted to measure the effect of increasing number of recommendation requests on the attained level of privacy. Figure 8 shows the privacy level while varying number of recommendation requests. In general, a decreasing number of recommendation requests decreases the privacy level, as the users will have a lower number of members in the formed peer-groups. However, the local concealment phase adds an extra layer of concealment to the released items' ratings which preserve a certain level of privacy for them.

In the sixth experiment, we wanted to measure the variation of the privacy levels as we increase the number of users in our system. We simulated a general case where there is one peer-group and the number of requests was fixed to be 12. Figure 9 shows the privacy level while varying the number of users participated within the various recommendation requests. With the increasing number of users, the density of users within peer-groups increases and thus each super-peer constructs a group profile for its peer-group which contains a considerable number of items' ratings. This in turn increases the privacy level for the data released by the super-peer for the PRS, as the super-peer has more potential to run an accurate global concealment on the group profile.

In the seventh experiment, we wanted to measure the variation of the privacy level as we increase the number of items' ratings sent within each request. Moreover, varying the number of items sent by each user has an effect on the traffic needs and work-load on the super-peer side as well as the accuracy of the predicated ratings. Figure 10, illustrates the privacy level while varying the number of items' ratings sent per each request. Increasing the number of items sent in each request increases the privacy level of the group profile, because the super-peer will have more scope to accurately perform the global concealment process.

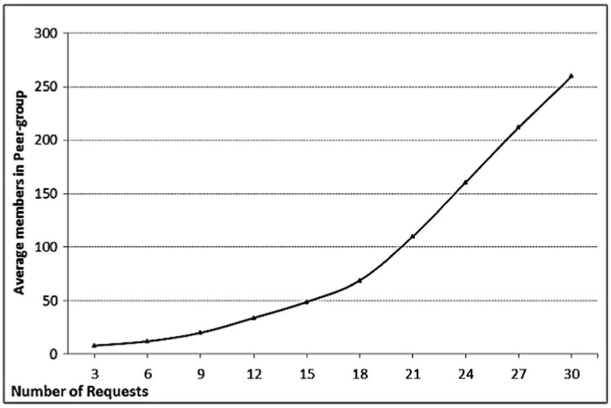


Fig. 7 Peer –group size as a function of the number of recommendation requests

6.1 Measuring the disclosure risk for the two stage concealment process

To measure the privacy or distortion achieved using CBO while realising the user's preferences to the super-peer, we will use the same metrics for privacy breach proposed in [19] to measure the true positive preferences that can be inferred from a user's profile α when he/she released this profile to the super-peer. Assuming this profile contains the preferences *s* which might be shared between all profiles released by other members of the peer-group. Based on this *precision*(s, a)=$|P_s^{shared} \cap P_a^{true}|P_s^{shared}$ will measure the portion of preferences that are shared by other members and they are truly consumed preferences for the user α and the *recall*(s,a)=$|P_s^{shared} \cap P_s^{true}|P_s^{true}$ refers to the portion of truly consumed preferences possessed by α that are actually in these shared preferences (privacy leak). A lower value for these metrics indicates a larger distortion between the shared and truly consumed preferences, which means a higher level of privacy achieved. In this experiment, we will evaluate the leaked private preferences of different users when running CBO. In this attack procedure, we will consider the users who published a portion of their truly consumed preferences in their released preferences, for each of these users; we tried to reveal other hidden preferences in their real profiles. The obtained preferences are quantified

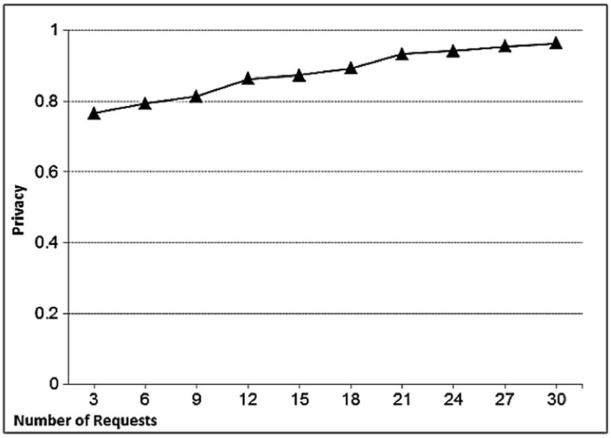


Fig. 8 Privacy level as a function of number of recommendation requests

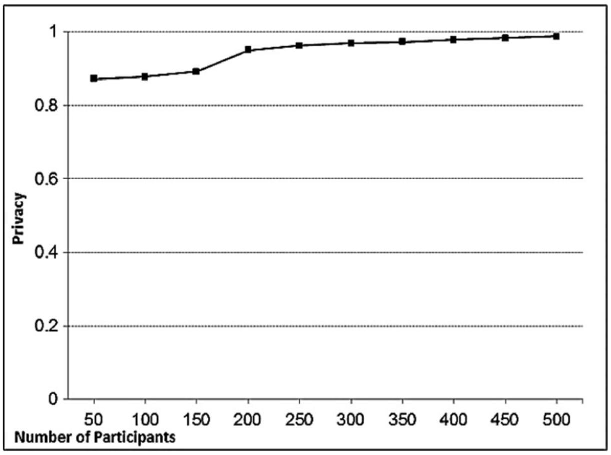


Fig. 9 Privacy level as a function of number of participants

using our proposed metrics and the results are shown in Fig. 11. As we can see, our CBO algorithm manages to reduce the privacy leakages for exposed users' true preferences. The privacy leak decreases as we increase the value of α (we only show for α values between 0.1 and 0.5). Based on this graph, we can state that the worst case for CBO prevent super-peers from inferring truly consumed preferences when α value is higher than 0.3.

However, our privacy metrics are pessimistic as the truly consumed preferences contained within the randomized portion published by the user and these preferences are already locally concealed at each user side. Moreover, the formation of peer-group mixes the collected preferences within a group profile. Therefore, such information disclosure has a limited impact on the preference breach.

In the final experiment, we wanted to measure the disclosure risk of our global concealment technique as the probability of reidentifying the globally concealed group profile based upon a portion of the originally released profiles from a malicious member within a peer-group. Inside this attack, a malicious PRS in collaboration with a malicious peer wants to filter out the

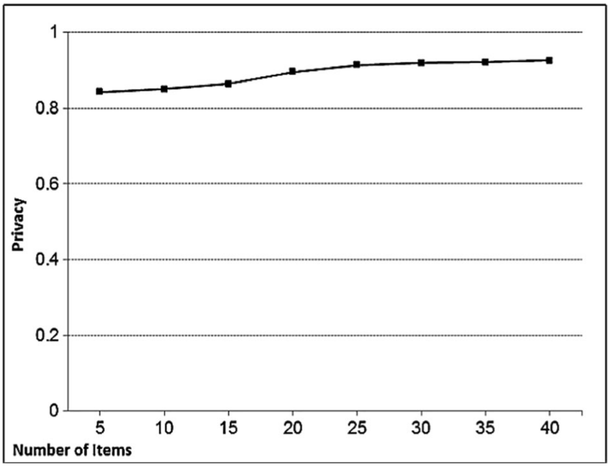


Fig. 10 Privacy level as a function of number of items

existing collected items from a group profile based upon a portion of the originally released items which have been disclosed by a malicious member within a peer-group in order to discover if certain preferences were released by the victim's profile. Due to that, we employed record linkage [10] metric to measure the difficulty of finding correct matches between the original preferences and its concealed version within the group profile. Given a group profile as the set $\gamma^o=\{\gamma_1^0,\gamma_2^0,\gamma_3^0\cdots\gamma_n^0\}$ he record linkage can be expressed as:

$$R_L=\frac{\sum_{i=1}^{n}Pr\left(\gamma_i^o\right)}{n}*100$$

Where $P_r(\gamma_i^o)$ is the probability of concealed rating for specific item, and it is computed as following:

$$Pr\left(\gamma_i^o\right)=\begin{cases}0 \;\; if \;\; \gamma_i^r \notin L\\ \frac{1}{|L|} \;\; if \;\; \gamma_i^r \in L\end{cases}$$

Where γ_i^r is the original rating, γ_i^o is its concealed version and L is the set of original ratings about a specific item, that has matched with the rating γ_i^o. Thus, we searched the original profile γ^r for matches with a concealed rating γ_i^o, the set of matched ratings L is searched for matches with γ_i^r based on an item. If $\gamma_i^r\in L$, then $P_r(\gamma_i^o)$ is calculated as the probability of finding γ_i^r in L. If no matches found, $P_r(\gamma_i^o)$=0. As shown in Fig. 12, we have computed $\mathrm{R_L}$ with respect to different values of ξ. As we can see, the lowest $\mathrm{R_L}$ is at ξ =10, the degree of $\mathrm{R_L}$ increases as ξ value increases. $\mathrm{R_L}$ is quite stable for ξ values between 40 and 50 and ξ =40 is considered as a critical point for $\mathrm{R_L}$. This allows the usage of higher ξ values while maintaining a sensible level of disclosure risk.

The threat model presented in Section 2, considers a malicious PRS who wants to discover the preferences within the user's profile by observing the items he/she releases for each recommendation request. The local concealment reduces this vulnerability by adding fake items similar to the real items in the released data in order to create a homogeneous concealed dataset. CBO within the local concealment selects a real item $\{I_u\}$ from real item set D_u with probability α or selects a fake item $\{I_u\}$ from the candidate fake item set D_n with probability $1-\alpha$. The selected item I_p is added

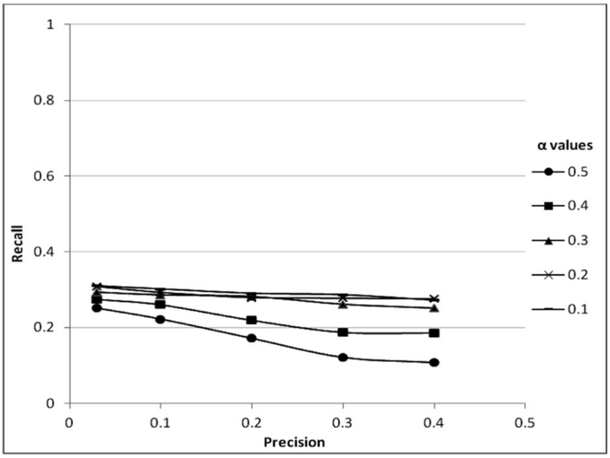


Fig. 11 The precision & recall for various values of α

as a record to the obfuscated dataset D_p. As a result, local concealment can reduce the leaked preference when releasing the data for different recommendation requests as the adversary cannot infer with confidence whether the item is in the participant's' profile or it is a fake item added to the released preference data. However, this does not come at the cost of accuracy as the fake items are semantically similar to the real items. Moreover, the formation of a peer-group mixes the collected preferences within a group profile. Trusted super-peers are elected based on their reputation, where each super-peer aggregates the locally concealed data obtained from the underlying users in a group profile and then it executes the global concealment using RRG algorithm by selecting a ξ percent of the partially unrated items to be randomly filling with a set of random values extracted from a distribution reflecting the ratings in the aggregated datasets. The globally concealed group profile is sent to PRS in order to perform rating predication in order to extract the referrals list. Therefore, any information disclosure will have a limited impact on the preferences breach.

7 Conclusion and future work

In this paper, we presented our ongoing work on building a collaborative privacy framework for preserving users' profile privacy in a social recommender service. We gave a brief overview of *EMCP* components and the interaction sequence for a recommendations process in an IPTV content distribution scenario. We presented a novel two stage concealment process that offers to the users a complete privacy control over their ratings profiles. The concealment process utilizes hierarchical topology, where users will be organized in peer-groups, from which super-peers are elected based on their reputation. Super-peers aggregate the preferences obtained from underlying users and then encapsulate them in a group profile and then send them to PRS. We tested the performance of the proposed framework on a real dataset. We evaluated how the overall accuracy of the recommendations depends on a number of users and requests. The experimental and analysis results showed that privacy increases under proposed middleware without hampering the accuracy of recommendations. Moreover, our approach reduces privacy breaches on the concealed data without severely affecting the accuracy of recommendations based on collaborative filtering techniques.

We realized that there are many challenges in building a collaborative privacy framework for preserving privacy in social recommender service. As a result we focused in middleware approach

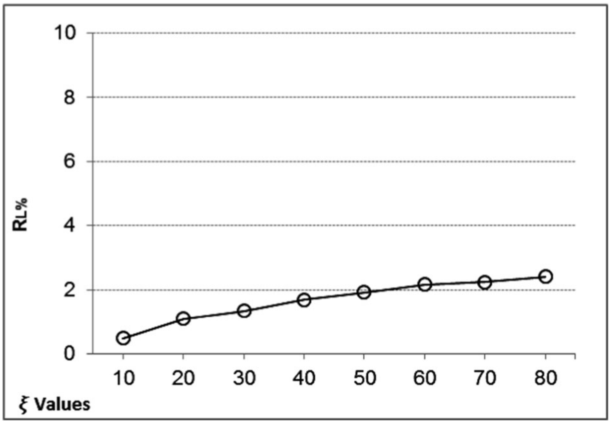


Fig. 12 R_L for the obfuscated dataset using RRG

for a collaborative privacy in an IPTV content distribution scenario. A future research agenda will include utilizing game theory to better formulate users groups, dynamic data release and its impact on privacy. Strengthen our middleware against shilling attacks and extending it to a p2p recommendation services. Moreover, we need to investigate weighed features' vector methods and its impact in released ratings. We need to perform extensive experiments in other real datasets from UCI repository and compare the performance with other techniques proposed in the literature. Finally we need to consider different data partitioning techniques as well as identify potential threats and add some protocols to ensure the privacy of the data against those threats.

Acknowledgments This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2013R1A1A2061978)

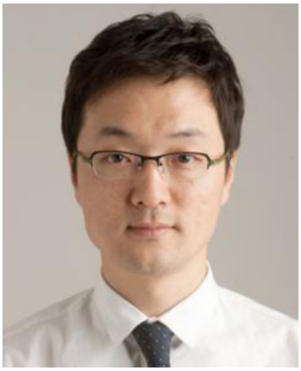
References

1. Ardissono L, Kobsa A, Maybury M (2004) Personalized digital television: targeting programs to individual viewers, vol 6, Human-Computer Interaction Series. Kluwer Academic Publishers, Boston
2. Beimel A, Nissim K, Omri E (2011) Distributed private data analysis: on simultaneously solving How and What. arXiv preprint arXiv:1103.2626
3. Blum A, Dwork C, McSherry F, Nissim K (2005) Practical privacy: the SuLQ framework. Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems. ACM, 128–138
4. Canny J (2002) Collaborative filtering with privacy via factor analysis. Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval. ACM, Tampere, pp 238–245
5. Canny J (2002) Collaborative filtering with privacy. Proceedings of the 2002 I.E. symposium on security and privacy. IEEE Computer Society 45
6. Carbo J, Molina J, Davila J (2002) Trust management through fuzzy reputation. Int J Coop Inf Syst 12:135–155
7. Commission E (2002) Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector. Official Journal L 201: 07
8. Cranor LF (2003) 'I didn't buy it for myself' privacy and ecommerce personalization. Proceedings of the 2003 ACM workshop on privacy in the electronic society. ACM, Washington
9. Directive E (1995) 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal of the EC 23: 6
10. Domingo-Ferrer J (2009) Record linkage. In: Liu L, Özsu MT (eds) Encyclopedia of database systems. Springer, US, pp 2353–2354
11. Dwork C (2006) Differential privacy. Automata, languages and programming. Springer, New York, pp 1–12
12. Eichhorn BH, Hayre LS (1983) Scrambled randomized response methods for obtaining sensitive quantitative data. J Stat Plan Infer 7:307–316
13. Elmisery AM, Botvich D (2011) An agent based middleware for privacy aware recommender systems in IPTV networks. In: Watada J, Phillips-Wren G, Jain LC, Howlett RJ (eds) Intelligent decision technologies, vol 10. Springer, Berlin, pp 821–832
14. Elmisery A, Botvich D (2011) Private recommendation service for IPTV system. 12th IFIP/IEEE international symposium on integrated network management. IEEE, Dublin
15. Elmisery A, Botvich D (2011) Agent based middleware for maintaining user privacy in IPTV recommender services. 3rd international ICST conference on security and privacy in mobile information and communication systems. ICST, Aalborg
16. Elmisery A, Botvich D (2011) Privacy aware obfuscation middleware for mobile jukebox recommender services. The 11th IFIP conference on e-Business, e-Service, e-Society. IFIP, Kaunas
17. Elmisery A, Botvich D (2011) Privacy aware recommender service for IPTV networks. 5th FTRA/IEEE international conference on multimedia and ubiquitous engineering. IEEE, Crete
18. Elmisery A, Botvich D (2011) Agent based middleware for private data mashup in IPTV recommender services. 16th IEEE international workshop on computer aided modeling, analysis and design of communication links and networks. IEEE, Kyoto

19. Elmisery AM, Doolin K, Botvich D (2012) Privacy aware community based recommender service for conferences attendees. 16th international conference on knowledge-based and intelligent information & engineering systems, vol 243. Ios Press, San Sebastian, pp 519–531
20. Esma A (2008) Experimental demonstration of a hybrid privacy-preserving recommender system. In: Gilles B, Jose MF, Flavien Serge Mani O, Zbigniew R (eds.) Vol. 0 161–170
21. Gemmis MD, Iaquinta L, Lops P, Musto C, Narducci F, Semeraro G (2009) Preference learning in recommender systems. European conference on machine learning and principles and practice of knowledge discovery in databases (ECML/PKDD). ACM, Slovenia
22. Golbeck J, Hendler J (2006) FilmTrust: movie recommendations using trust in web-based social networks. Consumer communications and networking conference, 2006. CCNC 2006. 3rd IEEE, Vol. 1 282–286
23. Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluating collaborative filtering recommender systems. ACM Trans Inf Syst 22:5–53
24. Hong T, Tsamis D (2006) Use of knn for the netflix prize. http://www.stanford.edu/class/cs229/proj2006/HongTsamis-KNNForNetflix.pdf
25. Huang Z, Chen H, Zeng D (2004) Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. ACM Trans Inf Syst 22:116–142
26. Huang Z, Du W, Chen B (2005) Deriving private information from randomized data. Proceedings of the 2005 ACM SIGMOD international conference on management of data. ACM, Baltimore, pp 37–48
27. Indyk P, Motwani R (1998) Approximate nearest neighbors: towards removing the curse of dimensionality. Proceedings of the thirtieth annual ACM symposium on theory of computing. ACM, Dallas, pp 604–613
28. Jannach D, Zanker M, Felfernig A, Friedrich G (2010) Recommender systems: an introduction. Cambridge University Press, Cambridge
29. Jeckmans AJ, Beye M, Erkin Z, Hartel P, Lagendijk RL, Tang Q (2013) Privacy in recommender systems. Social media retrieval. Springer, New York, pp 263–281
30. Kargupta H, Datta S, Wang Q, Sivakumar K (2003) On the privacy preserving properties of random data perturbation techniques. Proceedings of the third IEEE international conference on data mining. IEEE Computer Society 99
31. Kawazoe K, Kakinuma R, Haneda Y, Minoura D, Minamoto S, Ishimoto H (2007) Platform application technology using the next generation network. Technical Review. NTT
32. Kelly D, Teevan J (2003) Implicit feedback for inferring user preference: a bibliography. SIGIR Forum 37:18–28
33. Konstan JA (2004) Introduction to recommender systems: algorithms and evaluation. ACM Trans Inf Syst (TOIS) 22:1–4
34. Lam S, Herlocker J (2006) MovieLens data sets. Department of Computer Science and Engineering at the University of Minnesota
35. Lewis DD (1998) Naive (Bayes) at forty: the independence assumption in information retrieval. proceedings of the 10th European conference on machine learning. Springer, New York, pp 4–15
36. Lin J-L, Cheng Y-W (2009) Privacy preserving itemset mining through noisy items. Expert Syst Appl 36:5711–5717
37. Lin J-L, Liu JY-C (2007) Privacy preserving itemset mining through fake transactions. Proceedings of the 2007 ACM symposium on applied computing. ACM, Seoul, pp 375–379
38. Margulis ST (2003) On the status and contribution of westin's and altman's theories of privacy. J Soc Issues 59:411–429
39. McSherry F, Mironov I (2009) Differentially private recommender systems: building privacy into the net. Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, Paris, pp 627–636
40. Miller BN, Konstan JA, Riedl J (2004) PocketLens: toward a personal recommender system. ACM Trans Inf Syst 22:437–476
41. Narayanan A, Shmatikov V (2008) Robust De-anonymization of large sparse datasets. Proceedings of the 2008 I.E. symposium on security and privacy. IEEE Computer Society
42. Nejdl W, Wolpers M, Siberski W, Schmitz C, Schlosser M, Brunkhorst I (2003) Super-peer-based routing and clustering strategies for RDF-based peer-to-peer networks. Proceedings of the 12th international conference on World Wide Web. ACM, Budapest, pp 536–543
43. Nissim K, Raskhodnikova S, Smith A (2007) Smooth sensitivity and sampling in private data analysis. Proceedings of the thirty-ninth annual ACM symposium on theory of computing. ACM, 75–84
44. Parameswaran R, Blough DM (2008) Privacy preserving data obfuscation for inherently clustered data. Int J Inf Comput Secur 2:4–26
45. Polat H, Du W (2003) Privacy-preserving collaborative filtering using randomized perturbation techniques. Proceedings of the third IEEE international conference on data mining. IEEE Computer Society625
46. Polat H, Du W (2005) SVD-based collaborative filtering with privacy. Proceedings of the 2005 ACM symposium on applied computing. ACM, Santa Fe, pp 791–795

47. Ricci F, Rokach L, Shapira B (2011) Introduction to recommender systems handbook. Recommender systems handbook. Springer, New York, pp 1–35
48. Shokri R, Pedarsani P, Theodorakopoulos G, Hubaux J-P (2009) Preserving privacy in collaborative filtering through distributed aggregation of offline profiles. Proceedings of the third ACM conference on recommender systems. ACM, 157–164
49. Thuraisingham B (2002) Data mining, national security, privacy and civil liberties. SIGKDD. Explor Newsl 4:1–5
50. Ziegler C-N, McNee SM, Konstan JA, Lausen G (2005) Improving recommendation lists through topic diversification. Proceedings of the 14th international conference on World Wide Web. ACM, Chiba, pp 22–32

Ahmed M. Elmisery is Currently a Lecturer in Computer Science, Technical College, Egypt and a Research Assistant at Telecommunications Software & Systems Group, Department of Computing, Mathematics& Physics at Waterford Institute of Technology (WIT),Ireland and. He received his B.Sc. degree in computer science from Faculty of Computer science, Mansoura University, Egypt (2001), and M.Sc in computer science from Arab Academy for Science & Technology, Egypt (2007). He has published over 28 research papers in national and international conferences. His research interests include security, cryptography, and machine learning. He is conducting research on in privacy & security for future telecommunication services. His work has been grounded to develop a privacy enhanced algorithms for outsourced data in healthcare systems and recommender systems scenarios.



Seungmin Rho received his M.S. and Ph. D Degrees in Computer Science from Ajou University, Korea in 2003 and 2008, respectively. He visited Multimedia Systems and Networking Lab. in Univ. of Texas at Dallas from Dec. 2003 to March 2004. Before he joined the Computer Sciences Department of Ajou University, he spent two years in industry. In 2008–2009, he was a Postdoctoral Research Fellow at the Computer Music Lab of the

School of Computer Science in Carnegie Mellon University. He had been working as a Research Professor at School of Electrical Engineering in Korea University during 2009–2011. In 2012, he was an assistant professor at Division of Information and Communication in Baekseok University. Now he is currently an assistant professor at Department of Multimedia at Sungkyul University in Korea. His current research interests include database, big data analysis, music retrieval, multimedia systems, machine learning, knowledge management as well as computational intelligence. He has published 100 papers in refereed journals and conference proceedings in these areas. He has been involved in more than 20 conferences and workshops as various chairs and more than 30 conferences/workshops as a program committee member. He has edited a number of international journal special issues as a guest editor, such as Multimedia Systems, Information Fusion, Engineering Applications of Artificial Intelligence, New Review of Hypermedia and Multimedia, Multimedia Tools and Applications, Personal and Ubiquitous Computing, Telecommunication Systems, Ad Hoc & Sensor Wireless Networks and etc. He has received a few awards including Who's Who in America, Who's Who in Science and Engineering, and Who's Who in the World in 2007 and 2008, respectively.

Dmitri Botvich received his Bachelor's and Ph.D. degrees in Mathematics from Moscow State University, Faculty of Mechanics and Mathematics, Russia, in 1980 and 1984, respectively. He is currently a Chief Scientist at the Telecommunication Software and Systems Group, Waterford Institute of Technology (Ireland). He led the PRTLI FutureComm project at the TSSG, and has coordinated and worked in a number of EU and Science Foundation Ireland projects. He has published over one hundred papers in conferences and journals, and currently supervises 7 Ph.D. students. His research interests include bio-inspired autonomic network management, security, trust management, wireless networking, queuing theory, optimization methods, and mathematical physics.