Privacy-Preserving Targeted-Advertising in Peer to Peer Networks

Preliminaries:

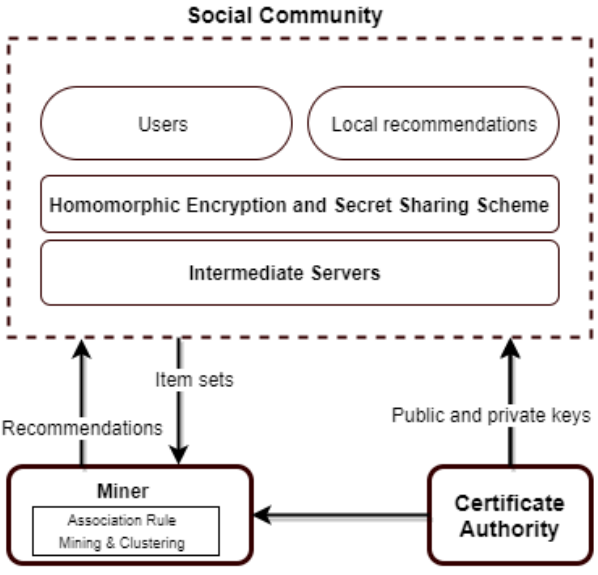
- Association Rule mining
  - Support
  - Confidence
- Homomorphic encryption, using
  - Elliptic curve cryptography
- Shamir's secret sharing
- K-means clustering

Design goals:

- The proposed system should be scalable and efficient
- No leakage of information at any point of the system i.e. privacy of all involving parties should be maintained throughout the design (No collusion of parties)
- Adversaries should not be able to affect the privacy and integrity of information passes through the communication channel
- Should be able to converge to a reasonable communication and computation cost
- Good recommendation accuracy

Methodology:

System overview:



To protect individual user's privacy, a system model is proposed. This system model, provides recommendations to the users without sacrificing the content interest to any party participating in the system. This system model targets large scale communities in the online social networks and is

designed to be scalable and efficient and at the same time providing privacy to the users. The proposed model consists of key components such as,

- **Online Social Community** on whom the targeted advertising is performed
- **Intermediate Servers (ISs)** interacts with the miner to obtain recommendations on behalf of real users
- **Miner** performs association rule mining on received item sets from users in online social community
- **Certificate Authority (CA)** distributes public and private keys to all the users to ensure secure transmission of item sets between the parties

1. User-Interest group definition and construction

This section, describes how user groups are organized in a privacy-preserving fashion.

Definition 1: A user group g is a two tuple: $\{U_g, I_g\}$ where $g \in G$ in which U_g is a set of users who share interests of the group they belong. Each Interest group I_g contains items of similar content. Users in the online social community are associated to the interest groups based on how similar items of users are to content of the interest group. User groups can be securely formed to hide privacy of each user from a set of users.

Next, we discuss how interest groups are formed and how users are associated to these interest groups. Before construction of user groups, identification of interest groups is most important as users are associated to these interest groups and user-interest based groups are formed. We divide the construction of user-interest based groups into three phases:

- Privacy-preserving user interest collection
- Identification of interest groups
- Association of users to the interest groups

1.1 Privacy-preserving user interest collection:

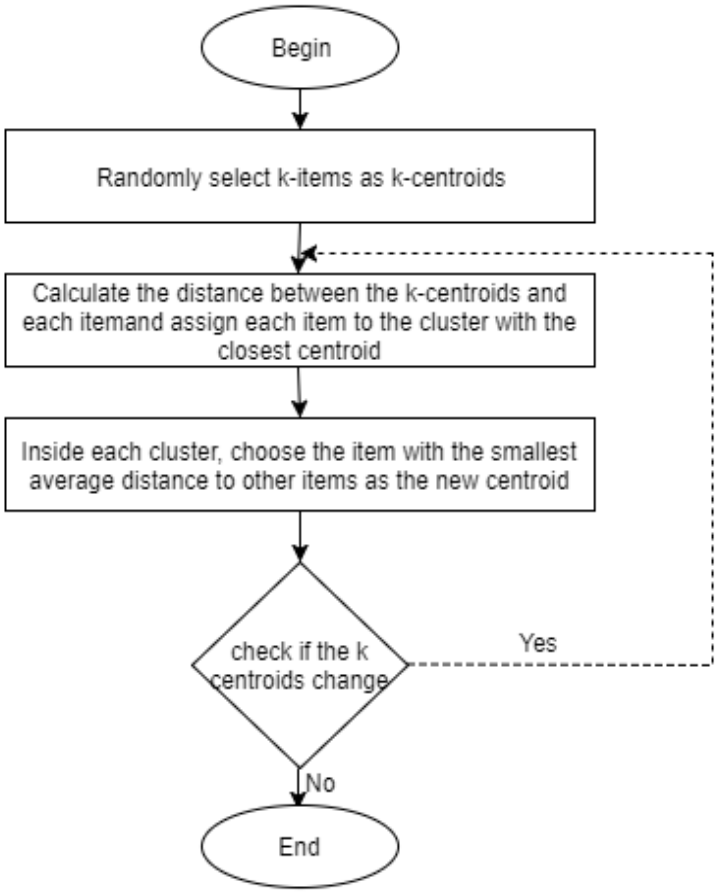
- Certificate Authority (CA) generates public and secret keys to all the users in social community, Intermediate-servers and the miner. CA distributes public keys of all users to other users and private key to respective users and a share of secret key of miner to ISs and miner.
- Users in the group calculate their local interests and encrypts the set with public key of miner and signs the encrypted item sets with its own secret key.
- Each user publish encrypted messages to respective pre-defined IS. ISs receives the signed interests from all the users and verifies the integrity and authenticity of signed messages with respective public keys of users.
- CA generates a polynomial, in which constant will be secret key of miner. Then it generates different shares of secret key of miner and distributes them to pseudo users and miner. Now each pseudo user has one share of secret key of miner.
- For reconstruction of secret key, miner needs shares from all the ISs to decrypt the interests of users.
- Once, miner reconstructs its secret key, can view all the interest sets of users which are signed by public key of miner by all users, without knowing to whom the interests belong.

Interest groups definition and construction:

This section describes how users are organized into groups in a privacy-preserving fashion. User group formation requires identification of interest groups and assigning users to interest groups. First, we define and identify interest groups securely without revealing items of users.

Definition 2: A set of interest groups $I_G = \{I_{g1}, I_{g2}, I_{g3}....I_{gk}\}$, where k is the number of interest groups, in which $I_g = \{i_1,i_2,i_3,....,i_m\}$ is a set of items and c_g belongs to I_g is the center of the group and represents "interest" of Ig and holds the property, for any two interest groups I_{gi} and I_{gj}, where $I \geq 1$, $j \leq k$ and $I \neq j$, I_{gi} intersection $I_{gj} = \emptyset$

It is important to identify true interests of users. Interest group identification ensures that users receive no "uninterested" items during recommendation process. For instance, in a community of "News", interest groups such as "sports news", "Business news", "Technology news" could be interest specific set of users, while a subcommunity could be interest to different set of users. In this model, we adopt k-centroid clustering algorithm to identify the interest groups which clusters similar items into interest groups. After interest groups are identified, each user will have an interest based on the items they like and similarity computation interest group they belong to.



Main challenges in identifying interest groups:

- Optimal number of interest groups i.e. good inter-group separation and intra-group similarity. A better number of interest groups helps to generate accurate recommendations to users
- Privacy-preserving item similarity computation

Association of users to Interest groups:

Given a set of users and the identified interest groups IG in the previous section, our goal is to associate users to the corresponding interest groups they belong. We consider the similarity function between the users and the interest groups. This function adequately captures the similarity of user's interest in different interest groups. This should be easy to calculate in a privacy-preserving fashion. Let U be the set of users and IG be the set of interest groups, we leverage item similarity(Ui,Ig) = |Ui intersection IG|/|Ui union IG|. Using this similarity function, interests of users with high similarity to the interest group are clustered into the same group.

Algorithm 1: ItemSimilarity(I_{gk},U)

Require: I_{gk} € I_G, i ≠ j,

1. ***for*** each u bel U ***do***
2. ***for*** each Ig bel I_G ***do***
3. *Compute* num = (u.I_g)
4. *Compute* den = (u*I_g)
5. **return** *similarity* = $\frac{num}{den}$
6. ***end for***
7. ***for*** each Ig with highest similarity ***do***
8. Assign user "u" to Interest group "I_g" with highest score
9. ***end for***

Given the similarity function, for a given k interest groups, miner can adopt k-centroids algorithm to cluster items into k different interest groups. Once items are clustered into k different interest groups, a user's interest determine the number of items he/she likes in each interest group and the group with highest similarity is the interest group to which the user belongs.

Privacy-preserving Recommendations:

After, identification of user-interest groups, miner can make recommendations at run-time to users, while protecting individual users interest information. To achieve privacy-preserving recommendations, the recommendation algorithm works in three phases:

- Identifying the interest group that an item to be recommended belong
- Aggregated and weighted ratings of each item from all users
- Personalized recommendations to the user, calculated on the client machine

Identifying the interest group an item belongs can help to calculate importance of each user to the interest group the item belongs. This can be determined using privacy-preserving Itemsimilarity function discussed in the previous sections. Personalized recommendations ensure good recommendation quality. Now we discuss, calculation of weighted ratings of each item within interest

group and followed by calculation of local personalized item rating of each user to achieve recommendation quality

Privacy-preserving item rating:

Let Ig be the set of interests in a group and Iu be set of items that user 'u' is interested in. Then, we define precision (u,g) and recall (u,g) of a user towards interest group as follows,

$$precision\ (u,g) = \frac{|Iu \cap Ig|}{|Ig|} \quad (1)$$

$$Recall\ (u,g) = \frac{|Iu \cap Ig|}{|Iu|} \quad (2)$$

Here, Precision (u,g) defines the fraction of items a user 'u' likes in the group 'g' and Recall(u,g) defines the fraction of item liked by 'u' are actually in 'g'. Both precision and recall are required to measure the importance of user 'u' to the group 'g

$$Weight\ (u,g) = \frac{2 \cdot Precision\ (u,g) \cdot Recall\ (u,g)}{Precision\ (u,g) + Recall\ (u,g\ )} \quad (3)$$

Weight (u, g) is computed and stored locally by each user which is later used to calculate personalized recommendations. After computation of individual users' importance to the group, we can also compute weighted rating of item 'i' in group 'g'. This is achieved by aggregating individual users rating of 'i', with privacy-preservation and the equation is as follows:

$$ItemRating(i)\ = \frac{\sum Weight(u,g) \times Ru,i}{\sum Weight(u,g)} \quad (4)$$

Here $R_{u,i}$ is the u's rating of item i, $R_{u,i}$ = 0 if u is not interested in item I, $R_{u,i}$ = 1 otherwise. More over to achieve personalized recommendations, each user finds group, the recommended item belong and Calculates the score using aggregated items ratings and weight of user to the group. Score is defined as,

$$Score_{u,i} = Weight(u,g)\ \times\ A_i \quad (5)$$

Where $A_{i,}$ is the set of aggregated ratings of all the items recommended, Weight (u,g) is the importance of the user in the group as defined above. Items with top $score_{u,i}$ are recommended to the user.