International Journal of Advanced Technology in Engineering and Science
Vol. No.5, Issue No. 05, May 2017
www.ijates.com
ijates
ISSN 2348 - 7550

ON ONLINE BEHAVIORAL BASED FRIEND RECOMMENDATION FRAMEWORKUSING SECRET SHARING SCHEME

Ketaki Handi[1], Sujata Jadhav[2], Manprit Kaur Kochhar[3], Sreelakshmi Hari[4]

[1,2,3,4]*Department of Computer Engineering, Pune University, Maharashtra (INDIA)*

ABSTRACT

Social network sites (SNS's) have connected millions of users creating the social revolution. The popular social networking sites are facebook, Twitter, MySpace, Orkut, LinkedIn, Google plus etc. which are actually online social networking (OSN) sites. a novel friend recommendation framework (FRF) based on the behavior of users on particular SNS's. The proposed method is consisted of the following stages: measuring the frequency of the activities done by the users and updating the dataset according to the activities, applying FP-Growth algorithm to classify the user behavior with some criteria, then apply multilayer thresholding for friend recommendation. The popularity of public cloud services, the concern of confidentiality is recognized as the problem even for personal individual users. Secret sharing data management approaches for multiple clouds to maintain confidentiality that involvesa secret sharing scheme have been proposed.

Keywords: Cloud computing, Recommendation Framework, Security, Social Networking Sites.

I. INTRODUCTION

Cloud computing has rapidly grown to be a platform of network-based computing. Public clouds have advantages in initial cost and availability. However, there are problems concerning confidentiality, such as improper use of data, because a third party's service provider manages data not only business user but also personal users. We focus on a storage service, which is a major cloud service. We have proposed a concrete data management approach. This approach uses secret sharing scheme. With our approach, confidential data are distributed to multiple cloud services using secret sharing scheme. In this paper, the proposed approach is implemented by using an actual cloud service as a CSP, and the performance is evaluated.

The popularity of online social networking sites is getting higher day by day because of the friendliness introduced in the sites and technological advancement.Friend recommendation is an important recommender application in social media. Major social websites such as Twitter and Facebook are all capable of recommending friends to individuals. However, most of these websites use simple friend recommendation algorithms such as similarity, popularity, or "friend's friends are friends", which are intuitive but consider few of the characteristics of the social network. A recommendation system generally interacts with its user in most possible friendly way and recommends doing something in its users favor. Recommendation systems for SNS's is a new scope of research as social peoples are more interested in online social networking (OSN) sites, like

Facebook, Twitter, Flickr, LinkedIn, MySpace, Google Plus etc. In the social networking sites, a social entity or user makes connections with other known or unknown social entities, namely friends or partners, and share their news and views through the profound facilities of the sites. Friends could be offline or real-life friends, classmates, neighbors, colleagues, family members, relatives or anyone having a profile in the OSN sites. Recommending people on social networking sites is worth studying because it is different from traditional recommendations of books, movies, restaurants, etc. due to the social implications of "friending". With the ever increasing web crimes and identity theft, people are becoming more and more careful in sharing their personal information. Hence, unless a user can trust the system with their data, the system cannot stand and it will be valueless. Exploitation of social network data is the fragmentation of the population of social network users into numerous proprietary and closed social networks. This issue is compounded by the fact that each new game or media application tends to build its own social network around it rather than building upon the rich data available about existing social relationships. In this paper, we have defined users' online behavior in perspective of mining and proposed a connection or friend recommendation framework for online SNS's using the concept of user's online behavior towards the other users.

II.SYSTEM ARCHITECTURE

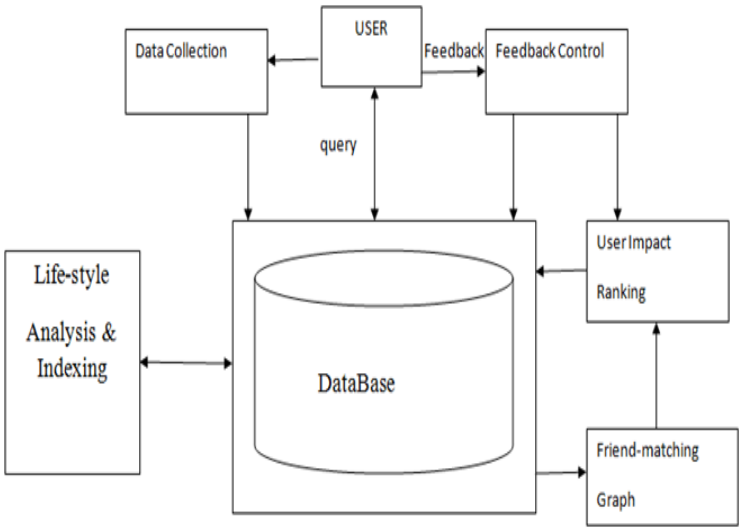

Fig1. System architecture

2 (a). Social Networking:

The popularity of online social networking sites is getting higher day by day because of the friendliness introduced in the sites and technological advancement. Use of these sites has developed social traditions and behavior in its users [1].

Nowadays, recommendation system has gained its popularity to the researchers' because of its versatile notion of integrating different research areas. Researchers from psychology, human computer interaction, computer vision, data mining etc. are keeping their attention on this research area. A recommendation system generally interacts with its user in most possible friendly way and recommends doing something in its users favor.

2 (b). Friend Recommendation Framework:

In this section, we proposed a novel friend recommendation framework based on the user's online behavior defined in the previous section. The framework isconsidered for a graph as the popular SNS's are architecture as graph based network. There are five step sequencing in the framework: extracting sub-network, finding frequency of the activities, find the common behavior, find the uncommonbehavior within the common behavior and finally friend recommendation.

- ***Extracting Sub-Network***

SNS's are very large entity and has large-scale databases.Day by day the size of the network is increasing and as the people are joining, there is huge number of information overload happens on these sites. For experiment of our proposed system, we take the whole network of a random individual. After getting the whole network of a client for who are going to suggest friends, we extract the sub network of 'x' no of people from the visualized graph.

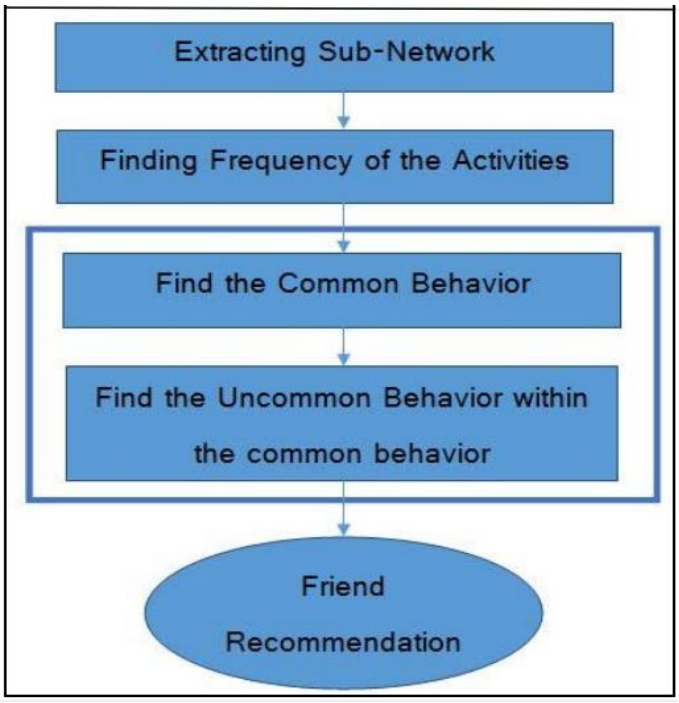

Fig 2: Proposed friend recommendation framework.

- ***Finding Frequency of Activities:***

Activities are the main base of social networking sites. The behavior of a particular user depends on the type of activities he/she does on the SNS. The definition of behavior also describes the related or considered activities actually define the behavior of that user. There are huge number of activities on SNS's nowadays and increasing day by day because of technological advancement and user's involvement from different spheres of life. For our proposed method we can consider different set of activities like types of songs user like, types of videos user often watches, types of online social games user take part etc. for each type of activities there will be many entities like a user may listen to "Michel Jackson" aswell as"Metallica". So there are two entities is listening

song activity.From all the activities we find out the entity with maximum frequency. We make the network scrutinized by using this frequency where the maximum frequency related activities are there. So all activities come down to the number of activities, where the every categorized activity contains one activity with the maximum frequency. After this step only the activities with maximum frequencies will be selected.

- ***Find the Behavior***

To find the desired user behavior (common and uncommon) we use FP-growth algorithm in our modified dataset. FP-growth algorithm gives us the pattern from the dataset. Among this pattern the desired behavior will be found. FP-Growth works in a divide and conquers way. It requires two scans on our model database. FP-Growth algorithm first computes a list of frequent items sorted by frequency in descending order (F-List) during its first database scan. In its second scan, the database is compressed into a FP-tree. Then FP-Growth starts to mine the FP-tree for each item whose support is larger than Į by recursively building its conditional FP-tree. The algorithm performs mining recursively on FP-tree. The problem of finding frequent item sets is converted to searching and constructing trees recursively. Thus, applying FP Growth algorithm in our dataset we find the behaviors patterns. From the acquired pattern we find out the common behavior by using the maximum number of frequency for any pattern with the single activities of the user behavior. Naturally single activities will give the highest frequencyvalue. After finding the common behavior our next step is to find the uncommon behavior. If we recommend friend only using the common behavior it will just like the natural recommendation process using only one feature like “fof” (friend of friend). So for more integrity we use uncommon behavior. This uncommon activity is different from the natural uncommon process. We actually find out the less no offrequency of other two activities from the user behavior containing the common user behavior. We call it uncommon because it appears less among the user whose have similar interest early. Actually this behavior can be considered as their unique behavior. Therefore, multilayer thresholding is done to find out the uncommon behaviors.

- ***Friend Recommendation***

In this final step we recommend the users with the user behavior found in previous steps. We can take any random user from any other sub network and recommend them as friend. So many friends or connections could be recommended to a particular user in any social network. The Fig-4 describes the recommendation framework more precisely. The total activity set is defined by the large circle here the inscribed circles define the different users and the activities they have performed in a particular timestamp.

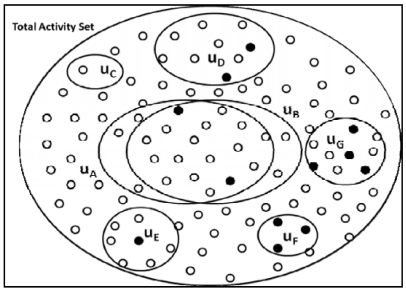

Fig 3: Behavior Analysis for friend recommendation forseveral users. (Black dots specifies the least frequentlyperformed activities and white dots specifies most frequentlyperformed activities)

Each user could have least frequently and most frequently performed activities which are denoted as black and white dots in the figure, respectively. Some users could have only most frequently done activities (i.e. uC), some could have only least frequently done activities (i.e. uF) or both (i.e. uA, uB, uD, uE, uG). Now, the uA and uB have many common activities as the sets overlaps and have two least frequently performed activities which is the commonly uncommon behavior that we have define in section III. Based on these activities these two users could be recommended to be friend to each other.

2 (c). Multiple clouds using secret sharing scheme

In the evaluation, a particular popular public storage service is used as a public cloud service because it is familiar and discloses its API to service developers [16]. Ideally, performance should be evaluated with many varieties ofcloud services, but here, one service is used with multiple accounts in order to clarify the basic characteristics because the effect on internet communication should be estimated first rather than the difference of various kinds of CSPs. Here, we use multiple accounts of the public cloud service as multiple CSPs. Evaluation with various services is future work.

- ***Evaluation Environment***

The configuration of the evaluation environment is depicted in Fig. 2. A PC for evaluation (Core i7, 2.8GHz) was connected through LAN (100 Mbps). A broadband connection for The Internet was used to communicate with the public cloud service as storage service. All programs were written in JAVA. AES was used as the encryption algorithm. The PC has functions of both client and DDS.

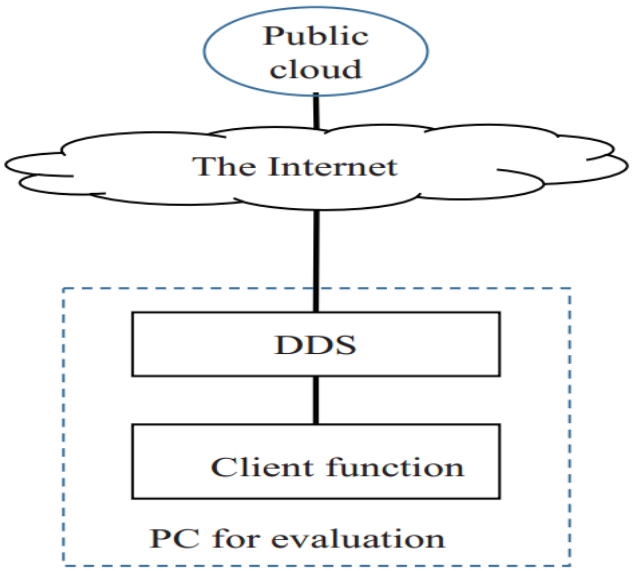

Fig 4. Evaluation Environment

Here, A fast (k, L, n) threshold secret sharing ramp scheme using XOR, proposed by, was used for secret share generation. The communication with the public storage service was performed by using disclosed API by the public storage service.

- ***Experimental Evaluation and Analysis***

Here, all measured point values are the value averaged by measuring 5 times. The results of evaluating pure processing performance where the parameters of the secret sharing scheme were fixed to k=3, L=2, n=5 in Fig. 3. Here, the secret data sizes were 1 and 10 MB. In Fig.4, process times of AES encryption/distribution comparing to process time of AES decryption/restoring are depicted except communication time to public cloud service. As shown in Fig.4, secret sharing process takes more time than AESprocess. Secret sharing process time is almost twice as long compared to AES process time, even in the case of encryption/distribution and decryption/restoration.

As shown in Fig.3, upload and download time with single account (no secret shearing) and the case that the secret sharing scheme were fixed to k = 3, L = 2, n = 5 were evaluated. In this case, multiple accesses (multiple accounts) perform in parallel. The upload uses 5 (n=5) accounts simultaneously, and the download uses 3 (k=3) accountssimultaneously. Both upload time and download time are nearly the same, though communication channels are varied. This is because communication speed of each channel is much lower than communication band and process time.

➢ *Basic Component Structure*

Figure 5 illustrates the basic logical component structure. We use multiple public clouds without adding extra sharing storages within an organization guaranteeing data security because we assume also easy personal usage. A client PC encrypts the secret data using a key. Usersthen transmit the encrypted data to the DDS (Data Distribution Server). The key is stored a client PC. The DDS applies the (k, L, n) secret sharing scheme to the encrypted data and transmits the generated share to each CSP. Furthermore, DDS generates a hash value to prevent falsification. Detailed processes are describes below.

➢ *Data Management Process*

The encryption/distribution and decryption/restoration with the secret sharing scheme are described.

○ *Encryption and Distribution*

1. A client who wants to store secret data M encrypts M with a key K and generates the encrypted data S. Then, the client discards M.
2. The client transmits S to the DDS. A secure protocol like SSL is necessary if the network is not trustable.
3. The DDS creates n shares DSj from S using the (k, L, n) secret sharing scheme.
4. The DDS generates hash values HDj from each DSj.
5. The DDS send hash value information and the identification information of shares to a client's storage.
6. The DDS distributes generated shares to multiple CSPs' storages.

- *Decryption and Restoration*

1. A client who uploaded S in the encryption and distribution procedure uploads HDj and identification information to the DDS.
2. The DDS selects arbitrary k cloud services out of n CSPs and requests the shares to each CSP, then, the DDS receives them.

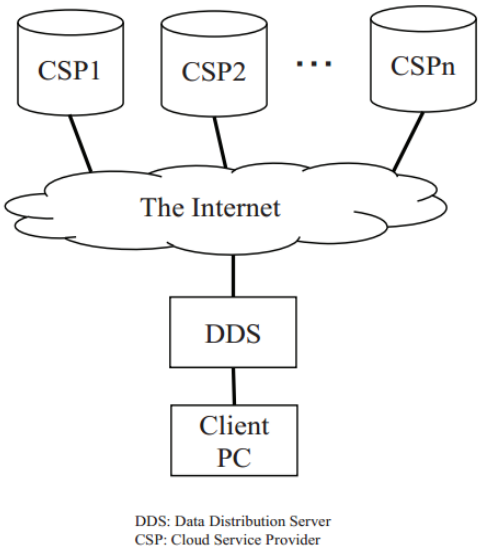

Fig 5. Assumed Basic Logical Component Structure

3. The DDS verify the received shares by using the hash values HDi, which are uploaded in Step 1. If the received DSi fails verification, the DS requests another share to another CSP.

4. The DDS restores S by applying the (k, L, n) secret sharing scheme.

5. A client receives S from the DDS.

6. The client decrypts S by using the key K and get original secret data M.

III. ALGORITHMS:

The following algorithms are used for friend recommendation and for security:

3 (a).FP-GROWTH:

To find the desired user behavior (common and uncommon) we use FP-growth algorithm in our modified dataset. FP-growth algorithm gives us the pattern from the dataset. FP-Growth works in a divide and conquers way. It requires two scans on our model database.

- FP-Growth algorithm first computes a list of frequent items sorted by frequency in descending order (F-List) during its first database scan.
- In its second scan, the database is compressed into a FP-tree.
- Then FP-Growth starts to mine the FP-tree for each item whose support is larger than α by recursively building its conditional FP-tree.
- The algorithm performs mining recursively on FP-tree. The problem of finding frequent item sets is converted to searching and constructing trees recursively.

Thus, applying FP Growth algorithm in our dataset we find the behaviors patterns. From the acquired pattern we find out the common behavior by using the maximum number of frequency for any pattern with the single activities of the user behavior. Naturally single activities will give the highest frequency value.

3 (b). AES:

AES is based on a design principle known as a substitution-permutation network, combination of both substitution and permutation, and is fast in software and hardware. Unlike its predecessor DES[20], AES does not use a Feistel network. AES is a variant of Rijindael which has a fixed block size of 128bits, and a key size of

128, 192, or 256 bits. By contrast, the Rijindael specification per se is specified with the block and key sizes that may be any multiple of 32bits, both with a minimum of 128 and a maximum of 256bits. Most AES calculations are done in a special finite field.

1. Key Expansion-

Round keys are derived from the cipher key using Rijindael's key schedule. AES requires a separate 128 bit round key block for each round plus one more.

2. InitialRound-

addRoundKey-

Each byte of the state is combined with a block of the round key using bitwise xor.

3. Rounds-

- Sub Bytes-

A non-linear substitution step where each byte is replaced with another according to a lookup table.

- Shift Rows-

A transposition step where the last three rows of the state are shifted cyclically a certain number of steps.

- MixColumns-

A mixing operation which operates on the columns of the state, combining the four bytes in each column.

- AddRoundKey.

4. Final Round(no MixColumns)

- SubBytes
- ShiftRows
- AddRoundKey

3 (c). DES:

DES relies on encryption techniques of confusion and diffusion.[20] Confusion means that each character of the cipher text should depend on several parts of the key. Diffusion means that a small change in the plaintext should reflect a big change in the cipher text, and similarly, a small change in the cipher text, should reflect a big change in the plain text. Confusion is accomplished through substitution, whereas diffusion is accomplished through permutation of the plaintext and the key. The algorithm includes initial permutation, sixteen identical iterations, 32-bit swap, and inverse initial permutation.[18] There is a key expansion function which produces different iteration keys that are used in each of sixteen iterations from the initial key.

The iteration function consists of the following operations:

• expansion permutation- in which the right half of the input block (Ri-1) is permuted and extended from 32 bit to 48 bit length,

• logical XOR function- which is applied to the extended right half of the input block and the iteration key for that iteration,

• substitution- which includes substitution function applied to the result of previous function which substitutes bits and reduces the size from 48 bit to 32 bit,

• permutation-which permutes the result from the previous function.

DES has a set of look up tables known as S-boxes and P-boxes, for substitution and transposition ofplaintext, respectively. Substitution algorithm uses eight S-box tables. Each of them gives 4-bit output for 6-bit input. Every S-box consists from four different 4-bit substitution algorithms. The most significant and the least significant bit from the input 6-bit value determines which of thefour algorithms will be used and the middle 4-bit value is replaced with the value from the table. After sixteeniterations and inverse initial permutation a 64-bit cipher text block is created. Iteration keys are created from 56-bit initial key. There are sixteen different 48-bit iteration keys. A permutation is applied to a 64-bit key that was entered by a user. This permutation also eliminates every eighth bit. Then there are sixteen iterations in which a 56-bit value is divided into two halves of equal size then for each half a circular left shift is done, followed by a permutation that selects 24 bit from each half. That is how a 48-bit iteration key is created. The result after rotation is input for next iteration.

IV. MATHEMATICAL MODEL:

Let S be the whole system consist of:

S = { U, w, z, d, Q, F}.

Where,

5. U is the set of number of users.

U = {u1,u2,…..un}.

6. Q is a set of query generated from user.

Q = {q1,q2,……qn}

7. F is the feedback of users.

F = {f1,f2,……fn}.

8. Let w is the set of activities.

w = [w1,w2,……wW]

wherewi is the ith activity and W is the total number of activities.

9. Let z is the set of life styles

z = [z1,z2,……,zZ]

Where zi is the ith life style and Z is the total number of life styles.

10. Let d is the set of life document.

d = [d1,d2,……, dn]

where di is the ith life document and n is the total number of users.

$$p(W_i|d_k)=\sum_{j=0}^{z} p(Wi|Zj)p(Zj|dk)$$

Let $p(W_i|d_k)$ is the probability of the activity $W_i$in a certain life document d_k.

Let $p(Wi|Zj)$ is the probability of how much the activity $W_i$contributes to the life style Zj.

Let $p(Zj|dk)$ is the probability of the life style Zj embedded in the life document dk.

V. CONCLUSION

We have proposed a novel friend or connection recommendation framework which could be used in any social networking sites. The framework is based on user's online behavior. In this paper, we have contributed the user's online behavior definition as well as an approach to use the online behavior to recommend friend. The

applications of this framework is huge and this approach could be used to recommend friend, community or group, online games matches with the users behavior or interest and many more. The FP Growth algorithm could be modified to determine a new recommendation system having more accuracy. Different data mining rules could be applied to simplify the model dataset and find the required connection. Our future work is to work with different data mining algorithms and large scale datasets from Facebook, Twitter, and MySpace etc.

We experimentally evaluated the performance of a proposed data management approach for multiple clouds that use secret sharing schemes by implementing the prototype. An actual particular public cloud service was used as a CSP in the prototype. The result shows that the performance wasfeasible for use and that the secret sharing processing time was much less than communication time. We will evaluate the performance with various kinds of CSPs in the future.

VI. REFERENCES

[1]. Friend Recommendation Framework for SocialNetworking Sites using User's Online BehaviorMd. Mehedi Hasan1, Noor HussainShaon 2, Ahmed Al Marouf 3, Md. Kamrul Hasan4, Hasan Mahmud5, and Md.Mohiuddin Khan6.

[2] S. Catanese, P. D. Meo, E. Ferrara, G. Fiumara, "Analyzing the Facebook Friendship Graph", 1st International Workshop on Mining the Future Internet, MIFI, 2010.

[3] L. Jin, Y. Chen, T. Wang, P. Hui, A. V. Visalakos, "Understanding User Behavior in Online Social Networks: A Survey", IEEE CommunicationMagazine, September 2013, pp. 144-150.

[4] X. Xie, "Potential Friend Recommendation in Online Social Network", IEEE/ACM International Conference on Green Computing and Communications, 2010.

[5] Y. Zheng, Y. Chen, X. Xie, W. Ma, "GeoLife2.0: A Location-Based Social Networking Service", 10th International Conference on Mobile Data Management: Systems, Services and Middleware (MDM), 18-20 May, 2009, pp. 357-358.

[6] A. A. Marouf, R. Ajwad, M. T. R. Kyser, "Community Recommendation Approach for Social Networking Sites based on Mining Rules", 2nd International Conference on Electrical Engineering
and Information & Communication Technology (iCEEiCT 2015), 21-23 May, 2015.

[7]. Performance Evaluation on Data Management Approach for Multiple Clouds Using Secret Sharing Scheme Atsushi KANAI, Shigeaki TANIMOTO and Hiroyuki SATO

[8]. J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy, "Make new friends but keep the old: Recommending people on social networking sites," in ACM CHI'09, New York, NY, USA, April 2009.

[9]. S. Wan, Y. Lan, J. Guo, C. Fan, and X. Cheng, "Informational friendship recommendation in social media," in ACM SIGIR'13, Dublin, Ireland, pp. 1045–1048, July 2013.

[10]. R. J. Rummel, The Conflict Helix: Principles and Practices of Interpersonal, Social, and International Conflict and Cooperation. New Brunswick, N.J, 1991.

[11]. M. Weber, The Nature of Social Action. Cambridge University Press, 1991.

[12]. H. Ma, H. Yang, M. R. Lyu, and I. King, "SOREC: Social recommendation using probabilistic matrix factorization," in ACM CIKM'08, Napa Valley, CA, USA, pp. 931–940, November 2008.

[13]. S. D. Roy, T. Mei, W. Zeng, and S. Li, "Socialtransfer: Cross-domain transfer learning from social streams for media applications," in ACM MM'12, Nara, Japan, pp. 649–658, October 2012.

[14]. N. Li and G. Chen, "Multi-layered friendship modeling for locationbased mobile social networks," in IEEE MobiQuitous'09, Toronto, Canada, pp. 1–10, July 2009.

[15] X. Xie, "Potential friend recommendation in online social networking," in IEEE/ACM CPSCom'10, Hangzhou, China, pp. 831–835, December2010.

[16]. V. Cardellini, V. Valerio, V. Grassi, S. Iannucci, F. Presti, "A New Approach to QoS Driven Service Selection in Service Oriented Architectures," Proc. of The 6th IEEE International Symposium on Service Oriented System Engineering SOSE, pp. 102-113, 2011.

[17]. G. R. Blakley, "Security of ramp schemes," Crypto'84, pp. 242-268,1984.

[18]. Pranav M, Archana K Rajan."DES security enhancement with dynamic permutation".

[19]. Zarko S. Stanisavljevic "Data Encryption Standard Visual Representation."

[20]. Wikipedia, Data Encryption Standard. https;//en.wikipedia.org/wiki/EFF_DES_cracker.