International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

Privacy Protection in Personalised Web Search

Ms. Rohini R Talekar[1], Dr. Sarika M Chavan[2]
[1, 2]*Computer Science and Engineering Department, Deogiri Institute of Engineering and management Studies, Aurangabad, Maharashtra India.*

Abstract: ***Nowadays everyone find information on internet. Whatever they want to search they search it on search engine. But not every time user gets information he is looking for, sometime user get information similar to his entered keyword. To solve this problem personalized web search is created, in personalized search user previous history is stored on server and by using that information user's profile is created. Now whenever user issues query the personalized search engine give them personalized result which is particularly made for that user. Personalized web search works very well when it comes to web results, but it has one problem. In PWS engine user information is saved on server for generating user profile, not every user is inclined to share his information. There is possibilities that this information is gets misused by adversaries, or any third party. In proposed method we tried to give solution to this problem by developing a client based server side personalized web search system. In proposed method we assumed that the server is trustworthy, but instead of just storing user data in its original form, we apply four different encryption techniques and algorithms on it. So when user issues query on this proposed system the query goes to server, server returns the results and store that query by applying algorithm on it such as, IB Encryption, Change Identity algorithm, Pollute Data algorithm, Handshake protocol. In proposed system we apply four times more security to user search history, so user can search information on search engine without any hesitation.***

I. INTRODUCTION

Privacy concerns have become the major problem for wide use of PWS services. The main goal of this project is not personalized search but privacy protection of user information which is getting stored on server, for protecting user data in profile-based methods; we have to consider two main points during the search process. On one hand, we have to increase personalization by using user's information. And on the other hand we have to hide that user profile information to give user more privacy. A few previous studies [10], [12] shows that users are ready to share their information with server if the output of personalization gives better results. In an ideal case, significant gain can be obtained by personalization at the expense of only a small (and less-sensitive) amount of the user profile. Thus, user privacy can be protected without compromising the personalized search quality. In general, there is a trade-off between the search quality and the level of privacy protection achieved from generalization method.

II. LITERATURE SURVEY

In this chapter we are reviewed all the existing system and related techniques of security in personalized web search.

A. Lidan Shou, He Bai, Ke Chen [1] uses a UPS framework to give privacy to user's hierarchical profile. Instead of creating user profile one time they use a online profile generalization. Their framework decides whether to personalize the query or not at online. Their framework works in procedures namely,1) offline profile construction, 2) offline privacy requirement customization, 3)online query topic mapping, 4) online generalization. In offline phase they build a user hierarchical profile which shows user interests, in second phase they ask user to mention sensitive node set, the nodes or keywords which user don't want to share with server. The purpose of query topping mapping is to make a rooted subtrees of *H* which is also called seed profile that contain all relevant topic in user issued query. In online generalization they build user runtime profile. For security they use two algorithms namely, GreedyDP and GreedyIL algorithm.in this UPS framework it works as a proxy on client machine before submitting user query to server they prune the sensitive node present in the issued query by using prune leaf technique. For calculating the risk of disclosing the sensitive node to server, they set one threshold value, so if during online generalization any sensitive node value is greater than threshold value that specific node is removed, thus user privacy is protected.

B. Single party private web search [2] also provide one solution to solve the privacy issue in personalized web search In this paper, the authors focus on those techniques that work directly in the computer of the users without requiring any external entity. More specifically, we propose a new single party scheme that addresses the trade-off between privacy and quality of service but it does not require any change at the server side. The performance of this new method has been evaluated using real search

queries extracted from the AOL's files. The results achieved show that their proposed method works as expected and it can be considered a proper option for those users who are concerned about their privacy. In this work, we use as knowledge base the Open Directory Project (ODP) hierarchy of categories because it is the largest, most comprehensive human-edited directory of the Web, constructed and maintained by a vast community of volunteer editors. In this project m fake queries are sent to user with original queries. ODP provides the list of m fake queries. This is done to prevent the search engine from ascertaining which one is the correct by observing their order of arrival. For example, let us assume that we want to obfuscate a one-word query like "tomato" considering m = 1. First, ODP retrieves its category "Top / Cooking / Soups and Stews / Fruit and Vegetable / Tomatoes". Then, ODP retrieves a fake category related to the original one; let us assume that this fake category is "Top / Cooking / Soups and Stews / Fruit and Vegetable / Carrots". From this category, ODP acquires the term "carrots" that is selected as our fake query. Finally, both queries are submitted to the WSE but only the search results linked to "tomato" are kept and presented to the user.

C. Ontology and hashing techniques [3] can also be used to provide security to user profile in PWS.in this when a user issues a query on the client for web search, it is encrypted using hash tag generated by Hashing technique. The encrypted query, which is in the form of a metadata, is passed to the server. The server then retrieves the date relevant to the query from the database after it has been processed by Taxonomy Management. The taxonomy management consists of content ontology and location ontology. The content ontology and the location ontology together with click through data are used to create feature vectors containing the user location preferences. They will then be changed into an area weight vector to rank the indexed lists as indicated by the client's location preference

III. SYSTEM DEVELOPMENT

The purpose of this chapter is to describe the overall design of proposed system of privacy preservation in personalized web search. It describes the architecture, its modules, its components and design of those components. Figure shows the architecture of proposed system

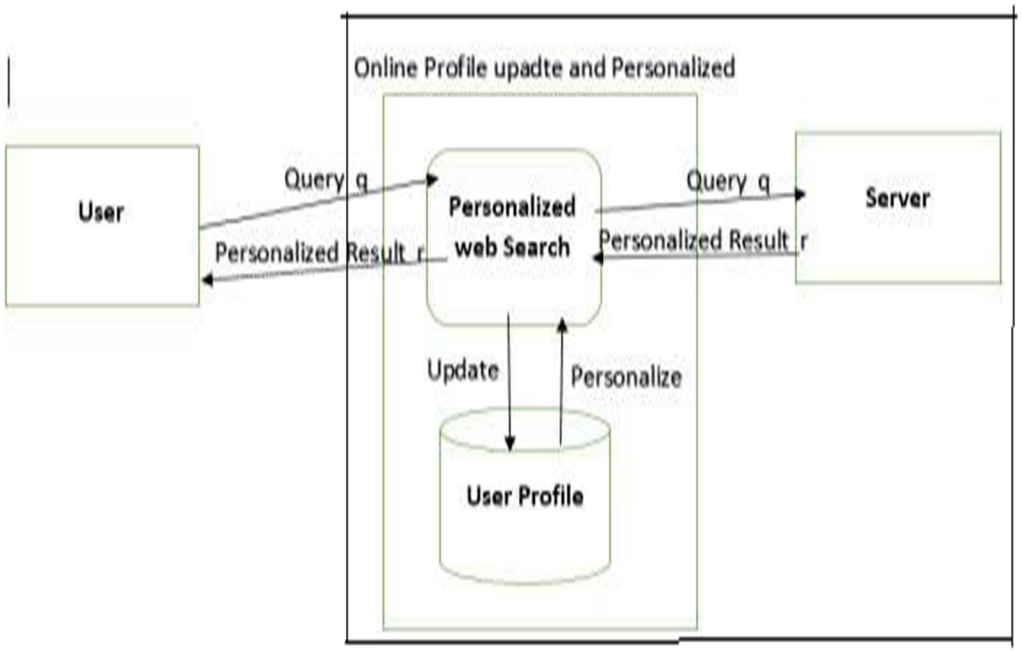

Fig 3.1: Architecture of proposed system

In proposed system we assume that the server is trustworthy for storing user profile information, but there is a possibility that user data may be leaked to adversaries, or misused so instead of storing original plain user query data of user we save it in encrypted format. Single encryption can't promise the total security of user search data and information; hence we used four different encryption algorithms and techniques on every single query issued by user before it gets stored on server database for PWS.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

The modules of project are as follows:

A. IB Encryption

ID-based encryption, or identity-based encryption (IBE), is an important primitive of ID-based cryptography. The first encryption algorithm we apply on user issued query is IB encryption. We use 128 bit key size to do encryption.

1) Protocol framework of IBE encryption

a) Setup: IBE run by private key generator (PKG). In this system server acts as PKG itself. This has all public and private key pair.

b) Encrypt: takes the message and output the encryption.

2) The logic used for encryption is as follows

1) Byte [] enc= (new BigInteger(m)).modpow(public_key,n).toByteArray();

2) Byte [] dec = (new BigInteger(enc)).modpow(Private_key,n).toByteArray();

3) System.out.println(new String(dec));

B. Change Identity Algorithm

After applying the IB encryption on users issued query we apply change Identity algorithm on it. In this algorithm any random user identity is picked to save in database instead of the current session user, thus the user identity is safe. While picking a random user this algorithm picks that user identity which has similar interest to as current user, thus his personalized web search can't be compromised. Once a random user is picked his identity is given to the current user only for one session. That same identity is not applied for every session.so even if third party get hold user's data they can't find the actual original user who issued the query because of the fake identity, thus user get personalized web search without revealing his true identity and his information and enhance his browsing experience.

C. Pollute Data

This is the third algorithm we apply on user entered query, in this noise is injected with user encrypted query.so even if during eavesdropping the attacker don't get the exact data of user. Here we used the server side privacy so when the random queries are put together, it chose those queries which is previously search on that PWS server, our noise injection model works as a black box with a selection switch inside. The black box generates diluted queries (Qs) by mixing noise queries (Qn) and user queries (Qu), then save that polluted queries on search engine.

This process continues until all Qu are sent. Thus Qs on the server side follows: $\forall i\ P\ (Qs = qi) = £P\ (Qu = qi) + (1 - £)P(Qn = qi)$

D. Handshake Protocol

After applying all above three techniques at last we apply handshake protocol to that noise injected query. In this Server has right to choose different encryption algorithm to encrypt that session of current user information. Different algorithms are selected for different session and different user.

Since the protocol is different for every session eavesdropper only get information of start and destination location they can't get real data.

IV. PERFORMANCE ANALYSIS

The proposed system assumes that server is trustworthy and we don't need user interaction for security, so PWS engine and server acts as a one module. Search engine process the request and display result, while server store user searched query in encrypted format.

Instead of using one algorithm we used four different algorithms. Single encryption can easily be breakable; hence we provide four times more security. Each technique overcome the drawback of other algorithm and balances them out, thus user get total security.

A. For Checking Personalization

For checking that our system gives personalized search results we create a database which has more than 100 url in it related to different topic. For adding websites we used crawler program. We use 20 different random key words to check the personalization of this system. As this is the small module it works fast and gives the personalized search results to each user. Following figure shows the personalization output of our system.

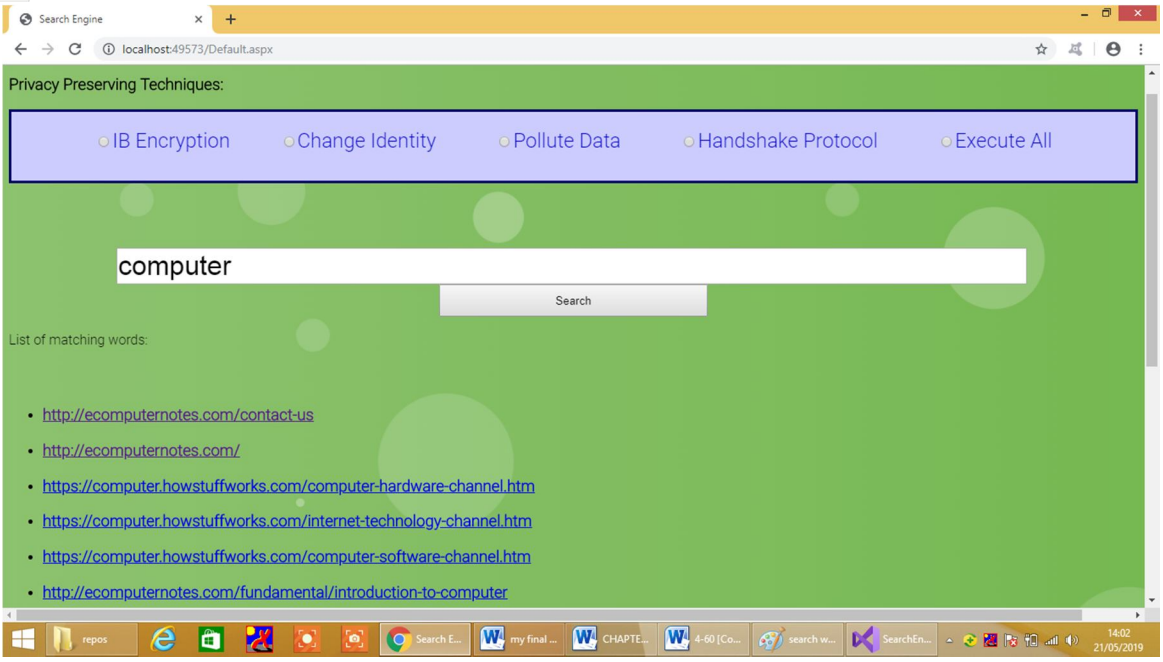

Fig 4.1: Output of Personalization.

B. For Checking Security

For checking security of our system with four techniques works together, we don't have any previous works which uses such kind of technique. Since this system uses four encryption and algorithms it's nearly unbreakable for adversaries or attacker to get the hold of user search history and profile.

If by any means they get user data, they have to first reverse the handshake protocol then they have to remove the noise we have injected with query, and then they have to find the real user identity and after that they have to decrypt the IB encryption. It states that it is nearly unbreakable to get exact information.

But personalization search and privacy are always inpropostional of each other. You cannot get both together. If lower the security faster the PWS result, and if higher the security slower the results.

Following fig shows the end to end time delay required by each technique individually to function.

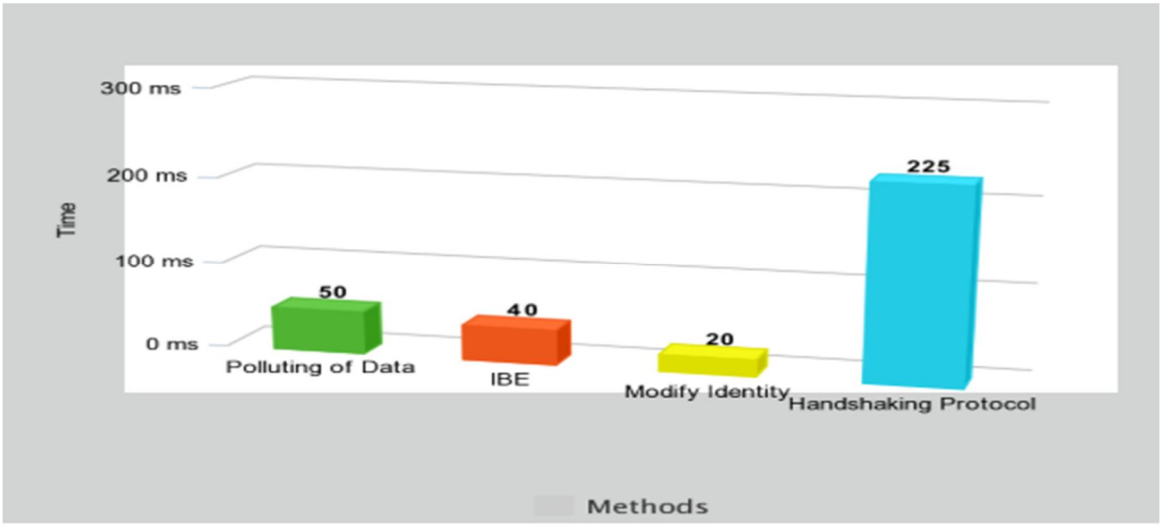

Fig 4.2: End to end delay of techniques individually

Since we applying all four together it require more time to present the results to user at expense of more security. Following graph shows the total time required by our system to encrypt queries.

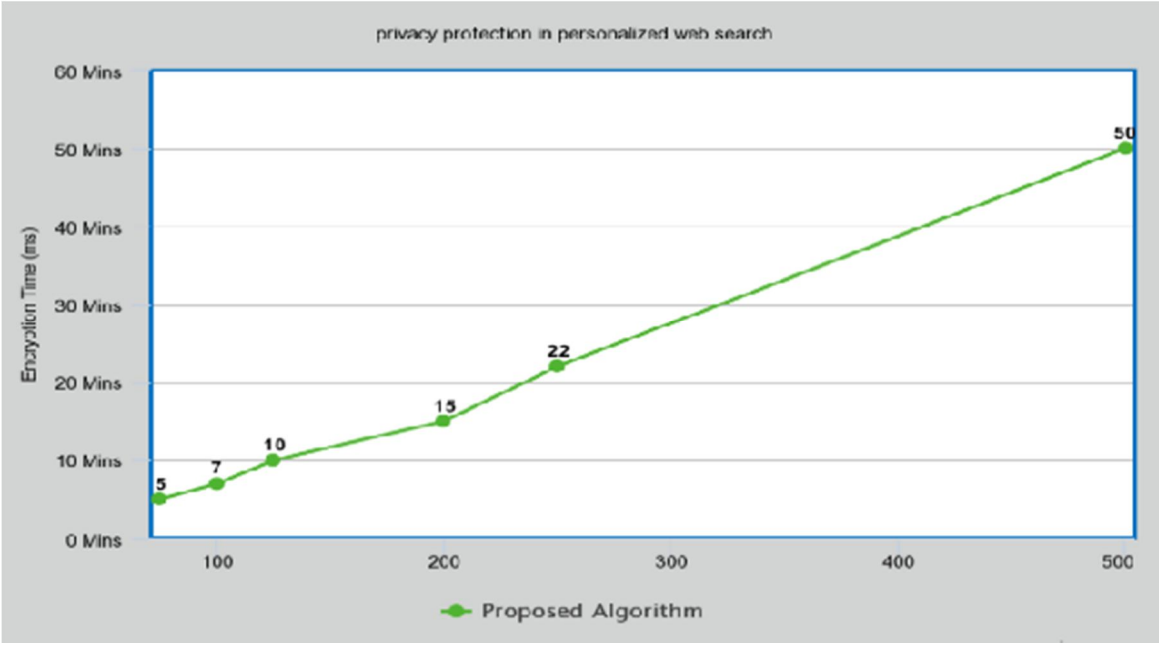

Fig 4.3: Encryption time required

You can't ever get faster search results and more security at the same time. They are always inversely proptional of each other. Following graph shows the comparison of existing system and proposed method. We compare our system with existing system which uses ontology and hashing techniques to provide privacy protection to user. In that they encryption techniques. We calculate the performance of both systems on the basis of time they required to encrypt and then display result. The x-axis on the figure shows the queries such as $20^{3,}$ 40^{3} etc. and y-axis states the time needed for encrypting that queries.

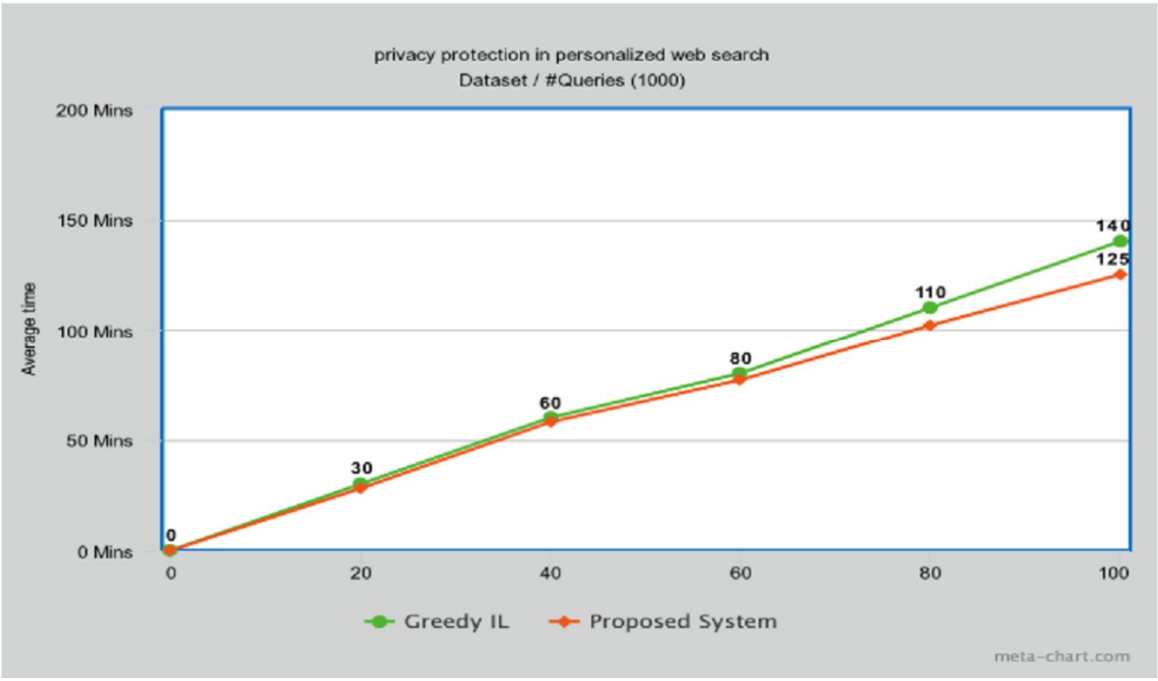

Figure 4.4: Performance analysis

At first proposed method does take same time as per existing system, but as the large number of queries increase our system works better than existing system with four times more security as an advantage. Following figure shows the output of system when applied four algorithms together.

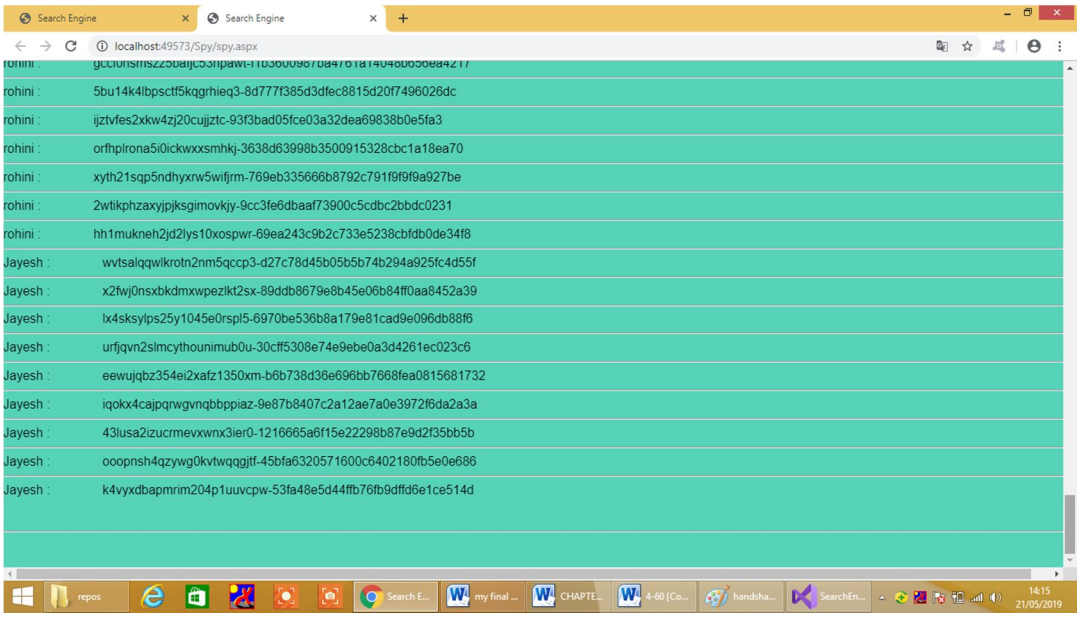

Fig 4.5: output window

V. CONCLUSION

We presented a novel system which helps to solve the problem of privacy protection in personalized web search. Here we used a client based server side search engine which provides personalized search results to each user specifically tailored for them by using their previous search history and user profile. To provide security to that user's search history we use four different algorithms which together encrypt that information before storing in the database. All four algorithms IB encryption, Change Identity, Pollute Data, Handshake Protocol are tested and analyzed for performance. In this proposed system we provided four layer securities to user profile to maximize the protection in personalized web search. This system can be adopted by any organization which have different branches situated physically at different places but share same database. It can also be used in medical related field. In proposed method we are promising user total security of his data thus, enhancing his browsing experience.

VI. FUTURE WORK

For future work, we are trying to increase the security by using safer and secure encryption algorithm. And also focus on increasing personalization utility of user information and minimize the time required for personalization results. The proposed method is well functioning for limited amount of queries, for future work we try to use larger and sophisticated server database which satisfy queries from large amount of clients.

REFERENCES

[1] Lidan Shou, He Bai, Ke Chen, Gang Chen "Supporting Privacy Protection in personalized web Search" proc IEEE /WIC vol -26 no-2 feb 2014.
[2] Alexandre Viejo, Jordi Castell`a-Roca, Oriol Bernad´o, Josep M. Mateo-Sanz " Single party private web search" published in conference in 2012 ISSN 236344568.
[3] Santosh S Kongbrailatpam, Dr. R.J. Anandhi "Securing personalised web search using ontology and hashing techniques" published in International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Vol 3 Issue 5 may 2015 ISSN no- 2320-9801
[4] Jaime Teevan Susan T. Dumais Eric Horvitz "Personalizing Search via Automated Analysis of Interests and Activities", International Journal of Advanced Research in Computer Science Engineering and Information Technology Volume: 2 Issue: -Mar-2014,ISSN_NO: 2321-3337
[5] Kumar, R.; Sharan, A., "Personalized web search using browsing history and domain knowledge," Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on , vol., no., pp.493,497, 7-8 Feb. 2014 doi: 10.1109/ICICICT.2014.6781332
[6] S. Ye, F.Wu, R. Pandey, and H. Chen, "Noise injection for search privacy protection," Department of Computer Science, University of California, Davis, Tech. Rep. CSE-2008-10, 2008.
[7] F. Saint-Jean, A. Johnson, D. Boneh, and J. Feigenbaum, "Private web search," in WPES'07: Proceedings of the ACM workshop on Privacy in electronic society, 2007, pp. 84–90
[8] G.-R. Xue, H.-J. Zeng, Z. Chen, Y. Yu, W.-Y. Ma, W. Xi, and W. Fan, "Optimizing web search using web click-through data," in CIKM'04: Proceedings of the 13th ACM international conference on Information and knowledge management, 2004, pp. 118–126.
[9] Handshake protocol information from https://en.wikipedia.org/wiki/TCP/IP Illustrated #Volume_1:_The_Protocols