ISSN: 2321-7782 (Online)
Volume 3, Issue 12, December 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Overview of a Security Problem in PWS

Upadhye Anirudha Pradip[1]
Department of Computer Engineering,
DGOI, FOE, Daund
Pune, Maharashtra-India.

Bere Sachin Sukhadeo[2]
Department of Computer Engineering,
DGOI, FOE, Daund
Pune, Maharashtra-India.

Abstract: ***Web search engines such as Google, Yahoo, and Microsoft Live Search are normally used to find certain data among huge amount of information in a negligible amount of time. Personalized search is a better way to improve the accuracy of web search, and it has been attracting much more attention in recent times. Efficient personalized search requires collecting and aggregating user information, which can raises serious worries of privacy violation for many users. However, users are not comfortable with showing theirs private information to search engines. Sometime privacy is not supreme, and often can be compromise if user gets gain in service or usefulness to the user. Thus, a balance must be struck between search quality and privacy protection [2].User's Profile provides a key input for performing personalized web search. In this paper, we exactly survey the issues of privacy preservation in personalized search. We differentiate and define four levels of privacy protection. We will also study that client-side personalization has benefits over the existing server-side personalized search services in preserving privacy. Here we are going to study details about open directory project.***

Keywords: data flow diagram; user customizable privacy-Preserving search; Open Directory Project Directory; User Profile.

I. INTRODUCTION

As the amount of information on the web constantly grows, it has become progressively tough for web search engines to find information that satisfies users' needs. Personalized search is a favorable way to improve search quality by customizing results for the people with different information goals. In practice, sometime, privacy is not absolute if people gain some economic benefits. One example is online shopping. They require to create profile and share personal information.**[2]**Web search engines help us to find out useful information on the Internet. There is lot of current research efforts have been focused on this area. Most of them could be categorized into two general approaches: Re-rank the query results returned by web search engines locally using personal information; or sending personal information and queries together to the search engine [1]. However, as the web corpus is on the server, re-ranking on the client side is bandwidth intensive because it requires a huge number of search results transmitted to the client before re-ranking. If the sum of information transmitted is restricted through filtering on the server side, it pins high trust on the survival of desired information among filtered results, which is not always the case. Though, the same query is submitted by different users, overall search engines return the same result without considering who submitted the query. Normally, each user has different information needs for his/her query. So the result should be displayed different for different user. There are three types of Web search systems which provide differentiated information:

1. Systems using relevance response,
2. Systems in which user register their information or interest, and
3. Users can give ratings, ratings on a scale from 1 (very bad) to 5 (very good).[3]

1) Inspiration

Search engine assist us to find any information on the web quickly and easily. But there is still opportunity for improvement. Existing web search engines do not consider specific needs of user and treat each user equally. It is difficult to let the search engine to know what the user actually want. There is one method which considers interest of each user and improves the quality of web search by suggesting the relevant pages of his/her interest. User privacy in profile-based PWS, researchers have to consider two opposing effects during the search process. On the one hand, they attempt to increase the search quality with personalization utility of the user profile. On the other hand, they want to hide from view the privacy contents existing in the user profile to place the privacy risk under control. Few previous studies have suggested that user is ready to compromise privacy if search engine returns good result. In an ideal case, considerable gain can be obtained by personalization at the expense of only a small and less sensitive portion of the user profile. In general there is a tradeoff between the search quality and the level of privacy protection achieved from personalization. Unfortunately previous works of privacy preserving PWS are far from optimal.

II. Literature Survey

1) Server Side Personalization

In this method personal information or user profile is stored on the search engine side. The search engine constructs and updates the user profile either through the user's explicit input (e.g., asking the user to specify his personal interests) or by gathering the user's search history implicitly (e.g., history, query and click through). Both methods require the user to create an account to identify user. Also, the client software requires no changes. This architecture is accepted by some search engines such as Google. Sometime server-side personalization architecture requires the user to give permission before his/her search history can be collected and used for personalization. If the user gives the permission, then search engine will collect all the personally identifiable information which is always available on the server side. Hence it even does not have level I privacy protection. It is not possible to implement Level III or IV privacy protection if personalization is done on the server side[4].

2) Client-Side Personalization

In client-side personalization every time identifiable information is stored on a user's personal computer. The user profile can be created from user requirement explicitly or search history implicitly. Client fires a query to the search engine and receives results, which is the same as in the general web search scenario. But when client sends query, a client-side personalized search system does query expansion to create a new query before sending the query to the search engine. It also re-ranks the search results. With this architecture, not only the user's search activities but also his related activities (e.g., web pages visited before) and personal information (e.g., emails, browser bookmarks) could be merged into the user profile, permitting for the construction of a much richer user model for personalization. The sensitive related information is generally not a major concern as it is strictly stored and used on the client side. Another advantage is that the overhead in computational storage can be distributed among the clients. A main disadvantage of personalization on the client side is, the personalization algorithm is not able to use some knowledge that is only available on the server side (e.g., PageRank)[4].

III. System Architecture of UPS

1) Data Flow Diagram

The data flow diagram (DFD) of UPS (user customizable privacy-Preserving search) is shown in fig1.System consist of a non-trusty search engine server and a number of clients. Basically each client is accessing the search service without trusting anyone. The most important component of privacy protection is an online profiler implemented as a search proxy running on the client machine. The proxy keeps both the complete user profile in a hierarchical form, and the user specific privacy requirements represented as a set of sensitive-nodes. This structure works on two phase, offline and online. In offline phase, a

hierarchical user profile is constructed and modified with the user-specific privacy requirements. And online phase works in following manner:

1. When a user issue a query, the proxy on the client generates new user profile in runtime.
2. The output of this step is generalized user profile, satisfying the privacy requirements of user.
3. The generalized user profile and the query are send together to the PWS server for personalize search.
4. Then results are personalized as per the users profile and send back to the query proxy.
5. Finally, the proxy presents the fresh results to the user or re-ranks result with the complete user profile.

UPS it provides runtime profiling, which in effect improve the personalization utility while respecting user privacy requirements. It also allows the customization of privacy need. Also it does not require iterative user interaction **[1].** This system considers user's profile (based on user's browsing history) and Domain Knowledge in order to perform personalized

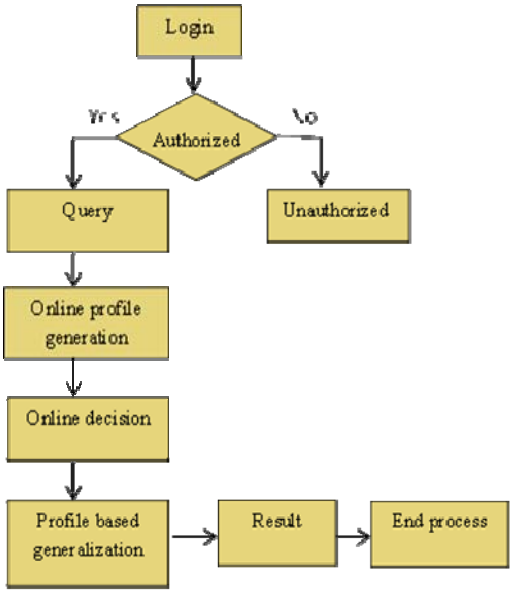

Fig. 1

web search. Domain Knowledge is used by the system store information about various domains/categories. This information is classified into these specified categories. Once the user inputs query, the system provides good recommendations for personalized web search based on user profile. It also re-rank the results obtain from the search engine based on the user profile.

2) Brief Overview of DMOZ Tool

The Open Directory Project, also named as the ODP or DMOZ, is the most important internet directory ever used. The DMOZ is the acronym for Directory Mozilla. The DMOZ listings are used by Google, and many other search engines. If you want to attain top rankings in the search engines, you must be included in the DMOZ Directory. Inclusion in DMOZ is required for high rankings in the Google search results as well. DMOZ is a completely volunteer edited directory, where every site is inspected by an editor for suitability. To use DMOZ directory you simply must get your site included in the ODP. The first step to a listing is to check the different directory categories, and find the proper one for your site. There exist a lot of categories, and sub-categories. If you submit your site to the incorrect category then you have to wait for a long time, and possibly even rejection.

Building a user profile for personalized web search is necessary. We can create a user profile in two ways. One is using commonly/frequently occurring words in user documents. This creates large user profiles where profile terms have low accuracy and have inadequate context to determine the user interests. The other method is by using a pre-existing ontology such as DMOZ. There is alternate method to build a hierarchical user profile using *Wikipedia*. It is created as the vocabulary for describing the user interests. The profile is created in this way are more compressed and have high precision as compared to profiles that use words. The ontology defines and restricts the vocabulary of the profile. The advantage of using DMOZ is that is an open-source, hence the responsibility of maintaining the ontology does not belong to one particular organization. The DMOZ tree is very large. It contains near about 600000 nodes. Personalization is playing an ever more important role in creating better internet search experiences. Current applications of personalization have focused on improving the search quality. An important feature of personalization is creation of a user profile. The user profile could be created on the client PC or on an Internet server. These both methods have different advantages and disadvantages. Client side profiles provide better privacy. Server side profiles enable collaborative filtering and profile portability. Most frequently occurring words in the user collection are use to represent the interest of user. Words at the top of the hierarchy represent overall interests of the user and words at the last represent specific interests. Disadvantage of using Wikipedia method is the large size of the Wikipedia index (about 1.4 GB). This restricts the use of the profiler on devices with low memory; we always try to work on reducing the size of the index. [5]

IV. PRIVACY PROTECTION IN PWS

In order to achieve better privacy in profile-based PWS, we have to consider two opposing effects during the search process. First is to improve the search quality. And second is hide the privacy contents existing in the user profile to keep the privacy risk under control. Several times people are ready to compromise privacy if search engine returns good result. The existing profile based PWS do not support runtime profiling. Existing system does not support ad hoc and ambiguous query. The existing system does not have the customization of privacy requirement system. So we have to consider two things: whether to personalize the query (by exposing the profile) and what to expose in the user profile at runtime [1].

1) Different Levels of Privacy Protection in Personalized Search

Each user has different requirements of privacy protection. Some users may not want anyone else to know their personal information, whereas others may be ready to share some personal information for getting better search results or services. Thus the level of privacy protection may need to be adjusted for different users according to their privacy requirement. In this section, we define and study four levels of privacy protection in personalized search

Level I – Pseudo Identity

Level I is the lowest level of privacy protection. Unluckily, this level is not sufficient to protect a user's privacy because it allows collection of all the information need descriptions of a user, which can in turn helps to identify the user.

Level II – Group Identity

Second level of protection is achieved when a group of users send their profiles to the search engine. With this the search engine can build a user profile for the group instead of creating for single user. In this personalization is possible at the group level. But it is not possible at the individual user level. This may decrease the effectiveness of personalization because a group's Information is used to model an individual user's information need. If the group is properly constructed so that people with similar interests are grouped together, we may have much richer user information. Because of the accessibility of more information from the group profile the search performance may be improved.

This level has higher privacy protection as compared to Level I. Here profile is created for a group of users. This level can be implemented using a proxy for a group of users and all the users will communicate with the search engine through the proxy. Presently, there are many public proxy servers present on the Internet. For example, TrackMeNot2, add on for Firefox web. .

Browser plug-in, which protects web searchers' identities by periodically issuing, randomized queries to search engines. Thus the aim of sending noisy queries is to make a single user profile seem like a group of user profiles, i.e., a user profile is not distinguish from a group of other user profiles.

Level III: No Identity

At this level a search engine cannot know ID of individual users. So there is no way to combine the description of user information needs. In this level, it's not possible to build a user profile on the search engine side. Personalized search must be supported on a user's own computer. To implement Level III privacy protection can be done through the anonymous network. For example, the web browser Torpark, it enables the user to communicate secretly on the Internet. When the user searches for a query using Torpark, the search engine would not be able to select where the search originally comes from. The search result is returned to the correct user through Tor network. At Level III, it is difficult to the search engine to aggregate any information about the individual user, even at the group level.

Level IV: No Personal Information

At this level, a search engine does not know identification of a user. Still, the search engine can return the normal search results to the correct user. Thus the user privacy is fully protected. At first sight, it appears to be impossible to achieve this level of privacy protection. At this level cryptography mechanism is used to achieve level IV protection. Another way is to use trusted third party. The search engine can release the index to a trusted third party. Initially user sends the query to the trusted third party and the trusted third party does the search for query and returns the results to the user. However, it is a challenge to design a communication protocol which give privacy guarantee on both the search engine side and the third party side. Since a search engine cannot construct any kind of user profile. Personalization is supported on the user's computer.

Another way to achieve this privacy protection is that a search engine would be required by law to guarantee that it does not store any user information. That is, the search engine will have no record of any activity of a user, even though it would still respond to a user search request directly. This situation can be considered to be equal to the scenario that the search engine doesn't know any information about the user. A search engine cannot build any kind of user profile; personalized search must be supported on the user's computer. In another form, the cost is that the search engine gives up the logging of any user information, which could be useful for other purposes such as anti-spam or detection of attacks. It has the uppermost level of privacy protection for personalized search. But, it may also have the maximum cost due to higher communication cost as well as encryption/decryption cost, which will delay real-time response.[**4**] In another form, the cost is that the search engine gives up the logging of any user information, which could otherwise be useful for other purposes such as anti-spam or detection of attacks.

V. CONCLUSION

Here we studied a client-side privacy protection framework called UPS for personalized web search. UPS can adopted by PWS that capture user profile in a hierarchical taxonomy. This framework allow user to specify his/her own customized privacy requirements via a hierarchical profiles [1]. UPS also performed online generalization on user profiles. It protects the personal privacy without compromising the search quality. We also studied privacy issues in PWS and different levels of privacy protection. We studied details about Open Directory Project called DMOZ. It is used to attain top rankings in the search engines. In future we will study various algorithms used in creating users profile. We will try to provide third and Fourth level security.

ACKNOWLEDGEMENT

I express great many thanks to Prof. Bere Sachin S. for their great effort of supervising and leading me, to accomplish this fine work. To college and department staff, they were a great source of support and encouragement. To my friends and family,

for their warm, kind encourages and loves. To every person gave us something too light my pathway, I thanks for believing in me.

References

1. "Supporting Privacy Protection in Personalized Web Search", Lidan Shou, He Bai, Ke Chen, and Gang Chen.
2. "Privacy-Enhancing Personalized Web Search", Yabo Xu* Simon Fraser University 8888 University Drive, Burnaby BC, Canada yxu@cs.sfu.ca .
3. "Adaptive Web Search Based on User Profile Constructed without Any Effort from Users". Kazunari Sugiyama Nara Institute of Science and Technology 89165 Takayama, Ikoma, Nara 6300192, Japan kazunas@is.naist.jp .
4. "Privacy Protection in Personalized Search", Xuehua Shen, Bin Tan, ChengXiang Zhai Department of Computer Science University of Illinois at Urbana-Champaign.
5. "Creating hierarchical user profiles using Wikipedia", Krishnan Ramanathan, Julien Giraudi, Ajay Gupta HP Laboratories HPL-2008-127.

AUTHOR(S) PROFILE

Mr. Upadhye Anirudha Pradip. Received his B.E. degree in Information Technology (Distinction) in the year 2012 from Solapur University. He is currently working toward the M.E. Degree in Computer Engineering from Savitribai Phule Pune University. He has 01 years of Industrial experience. His research interests lies in Data Mining.

Prof. Bere Sachin S. received his B.E. degree in Technology (First-class) in the year 2008 from Pune University and M.Tech Degree (Distinction) in Computer Engineering in2013 from JNTU University. He has 07 years of teaching experience at undergraduate and postgraduate level. Currently he is working as Assistant Professor and HOD in Department of Computer Engineering of DGOI, FOE, swami-chincholi, Daund, Pune University. His research paper has been published in IJTITCC, IJISET year 2014. His research interests are Digital Image processing.

