User customizable Privacy-preserving Search Framework-UPS for Personalized Web Search

Ms. A. S. Patil[1],Prof. M.M.Ghonge[2],Dr. M. V. Sarode[3]
Computer Science & Engg[1], Computer Engg[2,3]
M.E Scholar (CSE) [1], Assistant Professor [2],H.O.D[3]
JCET,Yavatmal- 445001, MS India[1, 2, 3]
Email: ankita.patil5@gmail.com[1]

Abstract-Searching is one of the common tasks performed on the Internet. Search engines are the basic tool of the internet, from where one can collect related information and searched according to the specified query or keyword given by the user, and are extremely popular for recursively used sites.The information on the web is growing dramatically. The users have to spend lots of time on the web finding the information they are interested in. Today, the traditional search engines do not give users enough personalized help but provide the user withlots of irrelevant information. In such case, personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' are not willing to disclose their private information during search has become a majorbarrier for the wide use of PWS. This paper gives information about privacy protection in PWS applications that model user preferences as hierarchicaluser profiles. This paper proposes a PWS framework called UPS that can adaptively generalize profiles by queries while respecting userspecified privacy requirements. It aims at providing protection against a typical model of privacy attack.

Index Terms-Privacy protection; personalized web search; UPS framework

1. INTRODUCTION

The web search engine is the most important portal for ordinary people looking for useful information on the web. However, users generally experience failure and get improperresults when search engines return irrelevant results that do not meet their real intentions. A typical search engineprovides similar set of results without considering of who submitted the query. Therefore, the requirement arises to have personalized web search system which gives outputs appropriate to the user as highly ranked pages. Personalized web search (PWS) is a general category of search techniques which aims to provide better search results, according to individual user needs. So, for this user information has to be collected and analyzed so that the perfect search results required for the user behind the issued query is to be given to the user. The solution to this isPersonalized Web Search(PWS). It can generally be categorized into two types, first is click-log-based methods and second is profile-based ones. The click-log based methods are simple and straightforward: This method performs the search based upon clicked pages in the user's query history. Although this method has been demonstrated to perform consistently and considerably well [2], it can only work on repeated queries from the same user, which is a strong limitation and restricted for certain applications. In contrast, profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. Profile-based methods can be proved more effective for almost all sorts of queries, but are reported to be improper under some situations.[1].Although there are reasons and considerations for both types of PWS techniques, the profile-based PWS has proved its more effectiveness in improving the quality of web search recently, with increasing usage of one'spersonal and behavioral information to profile its users, which is usually gatheredimplicitly with the help of query history [2], [3], [4], browsing history[5], [6], click-through data , [2] bookmarks, userdocuments [2], and so on Unfortunately, suchtype of collected personal data can easily reveal a entire scope of user's private life. Protecting privacy issues rising from the lack ofprotection for such data, for example the AOL query logsscandal, not only raise panic amongindividual users,but also downs the data-publisher's enthusiasm inoffering personalized service. In fact, privacy concernshave become the major barrier for wide use ofPWS services.

2. BACKGROUND

To protect user privacy in profile-based PWS, researchershave to consider two important and contradicting issues during the searchprocess. The first issue is that, they attempt to improve thesearch quality with the personalization utility of the userprofile. On the other hand the second issue is, they need to hide the privacycontents existing in the user profile to place the privacyriskunder control. Sometimes people are willing to compromise privacy ifthe personalizationby supplying user profile to the search engineyields better search quality. In an identical situation, significantgain can be obtained by personalization at theexpenseof only a small (and less-sensitive) portion of the userprofile, namely a generalized profile. Thus, user privacy canbe protected without compromising the personalizedsearch quality. In general, there is a compromise between thesearch quality and the level of privacy protection achievedfrom generalization.

Unfortunately, the previous works of privacy preserving PWS are far from optimal. The problems with the existing methods are explained in the following observations:[5]

1. The existing profile-based PWS do not support runtime profiling. A user profile is typically generalized for only once offline, and used to personalize all queriesfrom a same user indiscriminatingly. Such “oneprofile fits all” strategy certainly has drawbacksgiven the variety of queries. It is proved that Profile-based personalization maynot even help to improve the search quality forsome ad hoc queries, though exposing user profile toa server has put the user’s privacy at risk. A betterapproach is to make an online decision on:

a.whether to personalize the query (by exposingthe profile) and

b. what to expose in the user profile at runtime.

Until now no previous work hassupported such feature.

2. The existing methods do not take into account thecustomization of privacy requirements. This probablymakes some user privacy to be overprotected whileothers insufficiently protected. For example, inall the sensitive topics are detected using anabsolute metric called surprised based on theinformation theory, assuming that the interests withless user document support are more sensitive.

3. Many personalization techniques require iterative userinteractions when creating personalized search results.They usually refine the search results with somemetrics which require multiple user interactions,such as rank scoring, average rank [8], and so on.This paradigm is, however, infeasible for runtimeprofiling, as it will not only pose too much risk ofprivacy breach, but also demand prohibitive processingtime for profiling. Thus, we need predictivemetrics to measure the search qualityand breachrisk after personalization, without incurring iterativeuser interaction.

3. RELATED WORK

The above problems are explained in the UPS (which meansUser customizable Privacy-preserving Search) framework.[5]The framework assumes that the queries do not contain any sensitive information, and aims at protecting the privacy in individual user profiles while retaining their usefulness for PWS.

As given in Fig. 1, UPS consists of a nontrusty searchengine server and a number of clients. Each client (user) accessing the search service trusts no one but himself/ herself. The key component for privacy protection is an onlineprofiler implemented as a search proxy running on the client machine itself. The proxy maintains both the complete user profile, in a hierarchy of nodes withsemantics, and the user-specified (customized) privacy requirements representedas a set of sensitive-nodes.

The framework works in two phases, namely the offline and online phase, for each user. During the offline phase, a hierarchical user profile is constructed and customized with the user-specified privacy requirements. The online phase handles queries as follows:

1. When a user issues a query qi on the client, the proxy generates a user profile in runtime in the light of query terms. The output of this step is a generalized user profile Gi satisfying the privacy requirements. The generalization process is guided by considering two conflicting metrics, namely the personalization utility and the privacy risk, both defined for user profiles.

2. Subsequently, the query and the generalized user profile are sent together to the PWS server for personalized search.

3. The search results are personalized with the profile and delivered back to the query proxy.

4. Finally, the proxy either presents the raw results to the user, or reranks them with the complete user profile.

UPS is distinguished from conventional PWS in that it

1) provides runtime profiling, which in effect optimizes the personalization utility while respecting user's privacy requirements;
2) allows for customization of privacy needs; and
3) does not require iterative user interaction

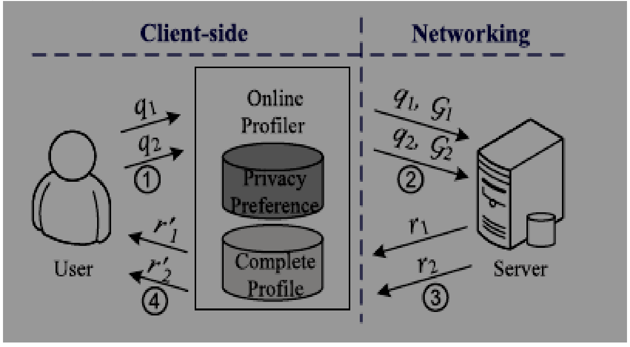

Fig.1. System architecture of UPS framework

.

4. ATTACK MODEL

Our work aims at providing protection against a typical model of privacy attack, namely eavesdropping. As shown in Fig. 2, to corrupt Alice's privacy, the eavesdropper Eve successfully intercepts the communication between Alice and the PWS-server via some measures, such as man-in-the middle attack, invading the server, and so on. Consequently,whenever Alice issues a query q, the entire copy of q together with a runtime profile G will be captured by Eve. Based on G, Eve will attempt to touch the sensitive nodes of Alice byrecovering the segments hidden from the original H and computing a confidence for each recoveredtopic, relying on the backgroundknowledge in the publicly available taxonomy repository R.

Note that in our attack model, Eve is considered as an adversary satisfying the following assumptions:[5]

Knowledge bounded: The background knowledge of the adversary is limited to the taxonomy repository R. Both the profile H and privacy are defined based on R.

Session bounded: None of previously captured informationis available for tracing the same victim in a longduration. In other words, the eavesdropping will be startedand ended within a single query session.

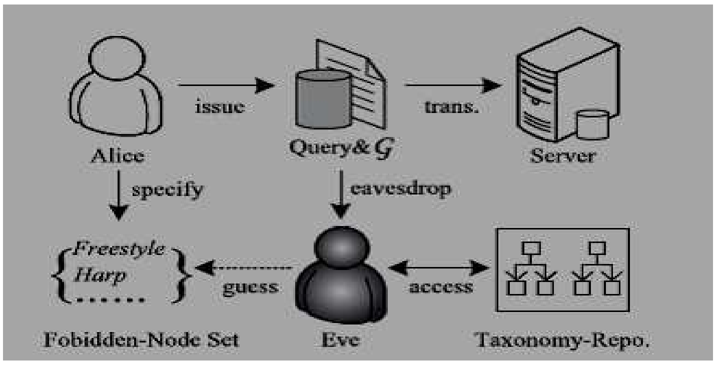

Fig. 2 Attack model representing personalized web search

5. CONCLUSION

The remarkable development of information onthe Web has forced new challenges for the construction of effective search engines. This paper provides information on User customizable Privacy preserving Search framework-UPS for Personalized Web Search. UPS could potentially be adopted by any PWS that captures user profiles in a hierarchical taxonomy. The framework allowed users to specify customized privacy requirements via thehierarchical profiles.

Acknowledgments

This work was guided by Prof. M.M. Ghonge and Dr. M. V. Sarode. I would like to thank both of them for their valuable guidance, help, and appreciate to work and constant source of inspiration.

REFERENCES

[1]B. Tan, X. Shen, and C. Zhai,(2006)Mining Long-Term Search Historyto Improve Search Accuracy, Proc. ACM SIGKDD Int'l Conf.Knowledge Discovery and Data Mining (KDD),.

[2] F. Qiu and J. Cho, (2006)Automatic Identification of User Interest for Personalized Search," Proc. 15th Int'l Conf. World Wide Web(WWW), pp. 727-736,

[3] J. Teevan, S.T. Dumais, and E. Horvitz, (2005)Personalizing Search viaAutomated Analysis of Interests and Activities, Proc. 28th Ann.Int'l ACM SIGIR Conf. Research and Development in InformationRetrieval (SIGIR), pp. 449-456

[4] K. Sugiyama, K. Hatano, and M. Yoshikawa,(2005)Adaptive WebSearch Based on User Profile Constructed without any Effortfrom Users, Proc. 13th Int'l Conf. World Wide Web (WWW),

[5]LidanShou, He Bai, Ke Chen, and Gang Chen (2014)Supporting Privacy Protection in Personalized Web Search, IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 2

[6] M. Spertta and S. Gach, (2005) Personalizing Search Based on UserSearch Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence
(WI),

[7]T.Sathiyabama, Dr. K. Vivekanandan (2011)Personalized Web Search Techniques -A ReviewByGlobal Journal of Computer Science and Technology Volume 11 Issue 12 Version 1.0 July

[8] X. Shen, B. Tan, and C. Zhai,(2005)Context-Sensitive InformationRetrieval Using Implicit Feedback, Proc. 28th Ann. Int'l ACMSIGIR Conf. Research and Development Information Retrieval (SIGIR),

[9] X. Shen, B. Tan, and C. Zhai, (2005)Implicit User Modeling forPersonalized Search, Proc. 14th ACM Int'l Conf. Information andKnowledge Management (CIKM),

[10]Z. Dou, R. Song, and J.-R.Wen,(2007)A Large-Scale Evaluation and Analysis of Personalized Search Strategies, Proc. Int'l Conf. WorldWide Web (WWW), pp. 581-590