International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Personalized Web Search Privacy Protection with Concealed User Requirement Specification

Devika P[1], Jitha K[2]

[1]PG Scholar, Department of CSE, MEA Engineering College, Malappuram, Under Kerala Technological University

[2]Assistant Professor, Department of CSE, MEA Engineering College, Malappuram, Under Kerala Technological University

Abstract: ***Personalized web search (PWS) verified its usefulness in refining quality search on internet. Unwillingness of users towards reveal their private data during search is the major obstruction for the extensive proliferation of PWS. This paper focus on replicas preference of users as hierarchical user profile. A framework with profile generalization and privacy preservation of user information keep the data secure. The run time generalization using two algorithms named GreedyDP and GreedyIL.***

Keywords: Profile personalization, Risk, Utility, Personalized Search

1. Introduction

Web search engine gained importance and lot of popularity for users data needs in web. The information accessible in web is ambiguous and vast. For the irrelevant queries by the user search engine results some failure. In order to get better search result PWS is used. By personalized web search the intension behind issued query can be examined. There it have two categories namely:
1) Click-log-based and
2) Profile based

The click log based methods are simple in nature and it simply examines the clicked pages of users query history. This method is followed for a period time with better result but problem in which the approach works on repeated queries leads to its limitation. This issue replaced with the help of profile based approach.

- The following are some of the problems listed by examining the existing methods.
- Run time profiling is not supported by the personalized web search.
- User profile used to personalize all queries from the same user and comprehensive only once offline.
- Search quality is not improved by profile based personalization.

In existing system sensitive issues are noticed by absolute metric called surprisal based on information theory which takes the interests with fewer user documents. The iterative user interactions are mandatory in personalization technique for creating personalized search results. Search results polished with some matrics such s rank scoring, average rank etc. This leads to difficulty for run time profiling. The risk factors increases and processing time increases.

2. Literature Survey

J.R Wen, R.Song and Z.Dou in [1] studied personalization on dissimilar uncertinity queries for different users under dissimilar background. Present an important valuation structure for personalized web search base on uncertain logs. Then estimate 5 personalized search approach utilize 10-day MSN uncertinity logs. The results are examined and exposed that personalized web search has important development over general web search on number of query, it also has tiny out come on additional question. Moreover it also demonstrates that simple click-based personalization method perform significantly and consconstantly, even as profile based are not under the proper work flow. Both long-term and short-term context are very significant in humanizing search performance for modified profile based approach.

G.Yang, K.Wang and Y.Xu in [2]introduced the notion of an anonymity in online to enable users to issue personalized queries for search to untrusted web service while their anonymity preserved. The challenge for providing is dealing with unknown and dynamic web users, who can get online and offline ay any time. Proposed the notion of online anatomy to ensure that each query entry in the query lig cannot linked to its sender, an algorithm that achieves online anatomy through user pool is introduced. So this approach can be extended up to deal with personally identifying information that may be contained in query. This method also applicable to general web based services where there is the need of anonymize the query, with or without personalization.

C.Verdery, Y.Zhu, and L.Xiong in [3] developed an optimal privacy notion to bound the prior and posterior probability that associate usr with an individual term in the anonymizing profile set of user was proposed. Authors introduced a novel bundling technique that can cluster users profile in to groups by taking account of semantic relationship between the terms while satisfying and following the privacy conditions. The problem of grouping user profile are studied and the goal is to prevent linking attacks that associate a user with individual term in the profile set that are anonymized.

J. Castellia-Roca And A. Viejo in [4] proposed new scheme which designed for protect the privacy of the users from a search engine that profile them. The distorted user profile to the web search engine are provided with the help of social networks. The standard queries are submitted to the search

engine, and does not require change in the server side. Server need not collaborate with users. The delay in query execution get reduced here. Proposed protocol preserve the privacy of individual deal with search engine. In this way queries get generated by user and submitted query group are not created.

Gang Chen and Lidan Shou in [5] examined hierarchical user structure for modeling user interest. The system provide generalization of user profile with online profiler at the client side. With the personalization utility system expected to enhance search efficiency, along with the privacy protection of user profile content.

3. Proposed System

This paper propose a user customizable privacy preserving framework. For every query as per user privacy specification get simplified by the use of the framework. Based on personalization and privacy risk metric it formulate generation of risk profile with demonstration of NP-hardness. The use of two Greedy algorithm named GreedyDP and GreedyIL for runtime generalization. The first one GreedyDP make the discriminating power and GreedyIL reduces the information loss. The mechanism for a client to choose a query and to personalize it with the framework is also considered. This decision making is completed before runtime profiling to enhance the stability of search result.

Advantages
1) The search quality steadiness get improved
2) The needless exposure of user profile avoided.
3) With the hierarchical profile allow customize needs.

Furthermore, User Customizable framework achieve online generalization on user profile look after personal privacy without compromising the quality of search.

As demonstrated in Figure 1,User customizable privacy preserving search framework contains no trusty search engine server and number of clients. Each and every client accessing the search service. The online profiler performed as a search proxy running on the complete client machine is the key module for privacy protection. The proxy uploads both the user specified privacy requirement signified as a set of sensitive nodes and the complete user profile in a hierarchy of nodes with semantics. The workflow carried out in two stages explicitly in the offline and online phase for every user. The offline phase hierarchical user profile is build and customized with privacy requirement. In online phase queries clutches as follows:

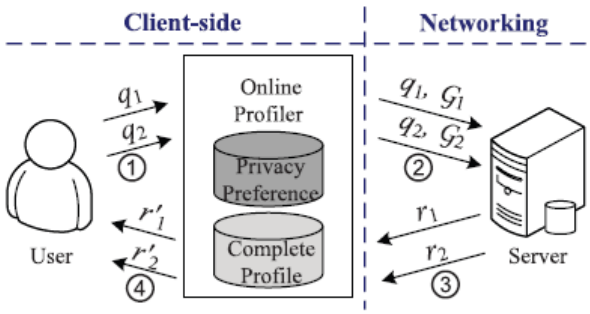


Figure 1: Framework Architecture

- User issues query q1 on the client, proxy create user profile in runtime, the light of query terms
- The outcome of the above step widespread user profile Gi sustaining privacy requirement.
- The query and user profile directed together to the PWS server, search result personalized with profile and return back to query proxy.
- Proxy either grants the raw result to user or rerank through complete user profile.

Greedy Algorithm
Greedy algorithm monitors the issue in resolving heuristic of creating the locally optimum choice at each stage with the confidence of finding a global optimum.It reflect easly to device and a simple approach.It decide next stage that provide advantageous result.In many problems greedy does not produce an optimal solution,but the greedy heuristic produces locally optimal solution that approximate a global optimal solution in sensible time.

i) GreedyDP:
- This algorithm works in bottom up manner. Starts from the leaf node for every iteration it choose leaf topic for pruning, trying to maximize utility of output.
- During each iteration a best profile-so-far maintained satisfying risk constraints, iteration stops when root topic reached, The final result is the best profile-so-far.
- Greedy DP algorithm requires recomputation of profiles which add up computational cost and memory requirement.

ii) GreedyIL:
- GreedyIL algorithm improve the generalization efficiency. It maintains a priority queue for candidate prune leaf operator in the decreasing order and reduces the computational cost.
- GreedyIL states to terminate the iteration when risk satisfied or when there is a single leaf left.
- Since there is less computational cost compared to GreedyDP, GreedyIL outperform GreedyDP.

Attack Model
The main goal of this work is focused in providing protection in privacy attack named eavesdropping. Here in figure 2 corrupt the Alice's privacy by a man in middle, that is Eve attempt communication between Alice and PWS server via some sort of techniques called man-in-the middle attack, invades the server. Similarly when Alice issue a query q, the full length copy of q along with a runtime profile G gets captured by Eve.

Based on G, Eve make an attempt to touch the sensitive nodes of Alice by recovering the segment hidden from the original H and every recovered topic, depending on the background knowledge in the publicly available taxonomy repository R.

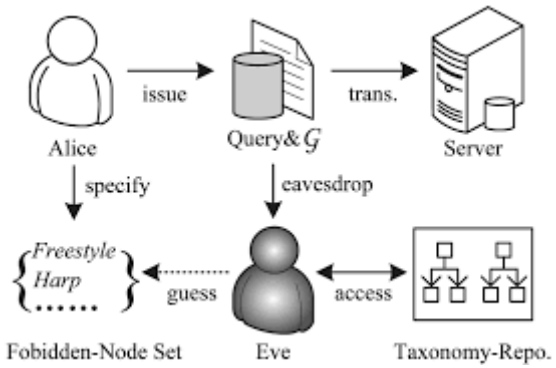
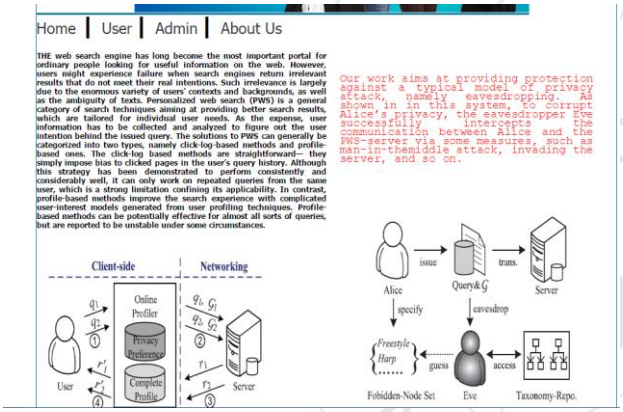


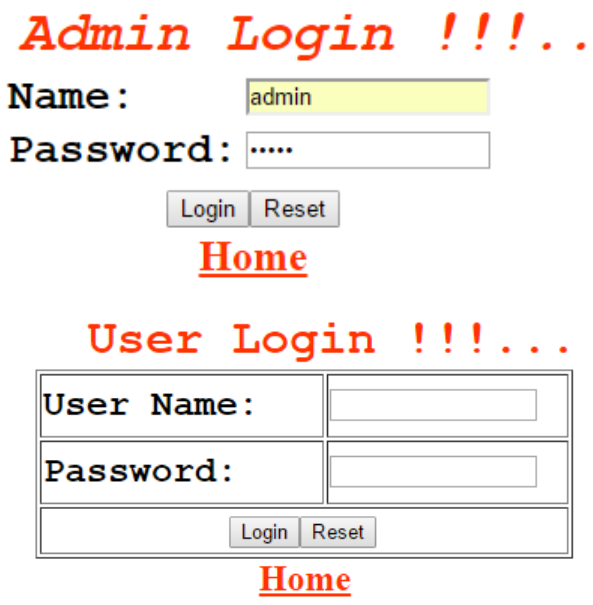
Figure 2: Eavesdropping

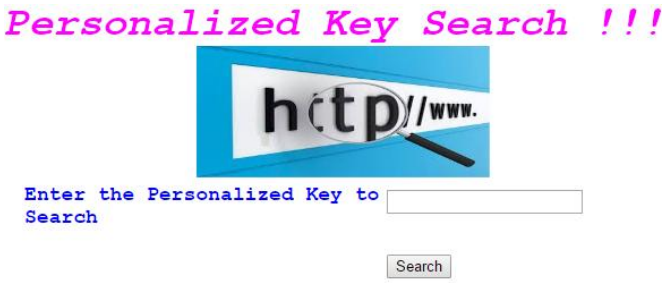
As discussed previously in our attack model, Eve is treated as an opponent sustaining the following assumptions:

Knowledge bounded: The background knowledge of the opponent is inadequate to the taxonomy repository R. So both the profile H and privacy are well defined based on R. Session bounded: Not any of previous captured information is presented for drawing the same victim in a long duration.

4. Result







5. Conclusion

This paper provides review on personalized web search and its security concepts.PWS technique are developed in last decades and variety of techniques emerged to improve search effectiveness and to protect privacy with multiple algorithms. Different method show that the privacy preservation is not in secure manner. The proposed framework provide privacy for user, uses online profiler to take online decision on a query to decide personalize it or not. It significantly reduce the of attack and perform better when compared to other. The main goal of this work is to assure privacy guarantee to the user in personalized web search.

Reference

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

[2] Y. Xu, K. Wang, G. Yang, and A.W.C Fu, "Online anonymity for personalized web services" *Proc.18th ACM conformation and knowledge management (CIKM)*, pp 1497-1500, 2009.

[3] Y. Zhu, L. Xiong and C. Verdery , "Anonymizing user profiles for personalized web search," *Proc.19th Int'l conf.World Wide Web(WWW)* , pp.1225-1226, 2010.

[4] Viejo and J. Castella-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines," Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.

[5] Lidan Shou, He Bai, Ke Chen, and Gang Chen "Supporting privacy protection in personalized web search" *IEEE Transactions on knowledge and data engineering*, Vol. 26, No. 2, February 2014.