ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (29-33), Month: April - June 2015, Available at: www.researchpublish.com

Supporting Confidentiality Protection in Personalized Web Search

[1] C. Indumathi, [2]R.Bharathi

[1]M.Tech (IT), Prist University, [2]Asst.Professor, Dept. of CSE, Prist University, Puducherry, India

Abstract: **Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. We study privacy protection in PWS applications that model user preferences as hierarchical user profiles. We propose a PWS framework called UPS that can adaptively generalize profiles by queries while respecting userspecified privacy requirements. Our runtime generalization aims at striking a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the generalized profile. We present two greedy algorithms, namely GreedyDP and GreedyIL, for runtime generalization. We also provide an online prediction mechanism for deciding whether personalizing a query is beneficial. Extensive experiments demonstrate the effectiveness of our framework. The experimental results also reveal that GreedyIL significantly outperforms GreedyDP in terms of efficiency.**

Keywords: **Privacy protection, personalized web search, utility, risk, profile.**

1. INTRODUCTION

Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. We study confidentiality protection in PWS applications that model user preferences as hierarchical user profiles.

We propose a PWS framework called UPS that can adaptively generalize profiles by queries while respecting user specified confidentiality requirements. Our runtime generalization aims at striking a balance between two predictive metrics that evaluate the utility of personalization.

We present two greedy algorithms, namely Greedy DP and GreedyIL, for runtime generalization. We also provide an online prediction mechanism for deciding whether personalizing a query is beneficial. Extensive experiments demonstrate the effectiveness of our framework. The experimental results also reveal that GreedyIL significantly outperforms Greedy DP in terms of efficiency.

2. EXISTING WORKS

A user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminatingly. Such "one profile fits all" strategy certainly has drawbacks given the variety of queries. One evidence reported in is that profile-based personalization may not even help to improve the search quality for some ad hoc queries, though exposing user profile to a server has put the user's confidentiality at risk.

The existing methods do not take into account the customization of confidentiality requirements. This probably makes some user confidentiality to be overprotected while others insufficiently protected. For example, in, all the sensitive topics are detected using an absolute metric called surprisal based on the information theory, assuming that the interests with less user document support are more sensitive. However, this assumption can be doubted with a simple counterexample: If a

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (29-33), Month: April - June 2015, Available at: www.researchpublish.com

user has a large number of documents about "sex," the surprisal of this topic may lead to a conclusion that "sex" is very general and not sensitive, despite the truth which is opposite. Unfortunately, few prior work can effectively address individual confidentiality needs during the generalization.

Many personalization techniques require iterative user interactions when creating personalized search results. They usually refine the search results with some metrics which require multiple user interactions, such as rank scoring, average rank, and so on. This paradigm is, however, infeasible for runtime profiling, as it will not only pose too much risk of confidentiality breach, but also demand prohibitive processing time for profiling. Thus, we need predictive metrics to measure the search quality and breach risk after personalization, without incurring iterative user interaction.

2.1 Disadvantages in existing Approach:

All the sensitive topics are detected using an absolute metric called surprisal based on the information theory. The existing profile-based Personalized Web Search do not support runtime profiling.The existing methods do not take into account the customization of confidentiality requirements. Many personalization techniques require iterative user interactions when creating personalized search results.

3. PROPOSED WORK

We propose a confidentiality-preserving personalized web search framework UPS, which can generalize profiles for each query according to user-specified confidentiality requirements. Relying on the definition of two conflicting metrics, namely personalization utility and confidentiality risk, for hierarchical user profile, we formulate the problem of confidentiality-preserving personalized search as Risk Profile Generalization, with its NP-hardness proved.

We develop two simple but effective generalization algorithms, Greedy DP and Greedy IL, to support runtime profiling. While the former tries to maximize the discriminating power (DP), the latter attempts to minimize the information loss (IL). By exploiting a number of heuristics, Greedy IL outperforms Greedy DP significantly.

We provide an inexpensive mechanism for the client to decide whether to personalize a query in UPS. This decision can be made before each runtime profiling to enhance the stability of the search results while avoid the unnecessary exposure of the profile

3.1 Advantages:

It enhances the stability of the search quality. It avoids the unnecessary exposure of the user profile.

4. MODULE DESCRIPTION

4.1 Profile-Based Personalization:

This paper introduces an approach to personalize digital multimedia content based on user profile information. For this, two main mechanisms were developed: a profile generator that automatically creates user profiles representing the user preferences, and a content-based recommendation algorithm that estimates the user's interest in unknown content by matching her profile to metadata descriptions of the content. Both features are integrated into a personalization system

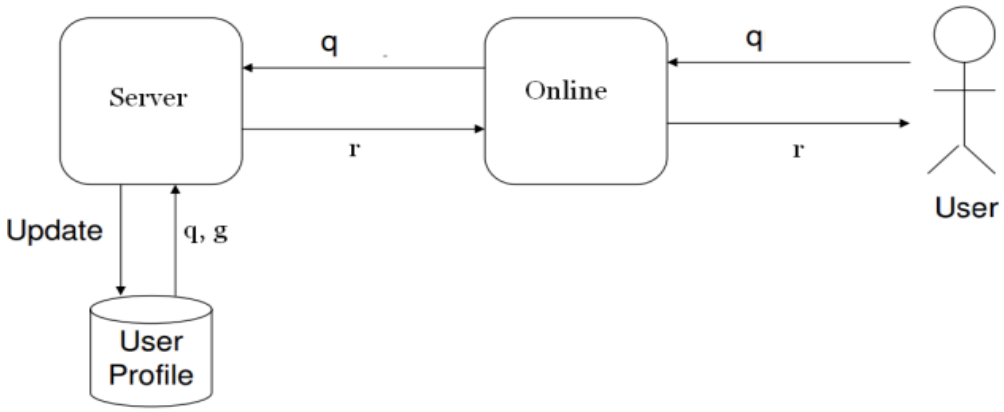

Fig.1.System architecture

4.2 Privacy Protection in PWS System:

We propose a PWS framework called UPS that can generalize profiles in for each query according to user-specified privacy requirements. Two predictive metrics are proposed to evaluate the privacy breach risk and the query utility for hierarchical user profile. We develop two simple but effective generalization algorithms for user profiles allowing for query-level customization using our proposed metrics. We also provide an online prediction mechanism based on query utility for deciding whether to personalize a query in UPS. Extensive experiments demonstrate the efficiency and effectiveness of our framework.

4.3 Generalizing User Profile:

The generalization process has to meet specific prerequisites to handle the user profile. This is achieved by preprocessing the user profile. At first, the process initializes the user profile by taking the indicated parent user profile into account. The process adds the inherited properties to the properties of the local user profile. Thereafter the process loads the data for the foreground and the background of the map according to the described selection in the user profile.

Additionally, using references enables caching and is helpful when considering an implementation in a production environment. The reference to the user profile can be used as an identifier for already processed user profiles. It allows performing the customization process once, but reusing the result multiple times. However, it has to be made sure, that an update of the user profile is also propagated to the generalization process. This requires specific update strategies, which check after a specific timeout or a specific event, if the user profile has not changed yet. Additionally, as the generalization process involves remote data services, which might be updated frequently, the cached generalization results might become outdated. Thus selecting a specific caching strategy requires careful analysis.

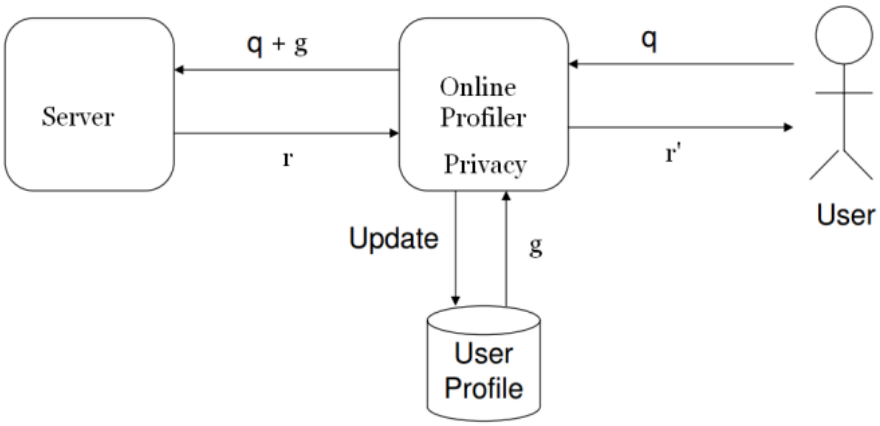

Fig.2. Enhanced Architecture

4.4 Online Decision:

The profile-based personalization contributes little or even reduces the search quality, while exposing the profile to a server would for sure risk the user's privacy. To address this problem, we develop an online mechanism to decide whether to personalize a query. The basic idea is straightforward. if a distinct query is identified during generalization, the entire runtime profiling will be aborted and the query will be sent to the server without a user profile.

5. PRELIMINARIES

5.1 User Profile:

Consistent with many previous works in personalized web services, each user profile in UPS adopts a hierarchical structure. Moreover, our profile is constructed based on the availability of a public accessible taxonomy, denoted as R, which satisfies the following assumption.

5.2 Customized Privacy Requirements:

Customized privacy requirements can be specified with a number of sensitive-nodes (topics) in the user profile, whose disclosure (to the server) introduces privacy risk to the user.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (29-33), Month: April - June 2015, Available at: www.researchpublish.com

5.3 Attack Model:

Our work aims at providing protection against a typical model of privacy attack, namely avesdropping. As shown in Fig. 3, 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 by recovering the segments hidden from the original H and computing a confidence for each recovered topic, relying on the background knowledge in the publicly available taxonomy repository R.

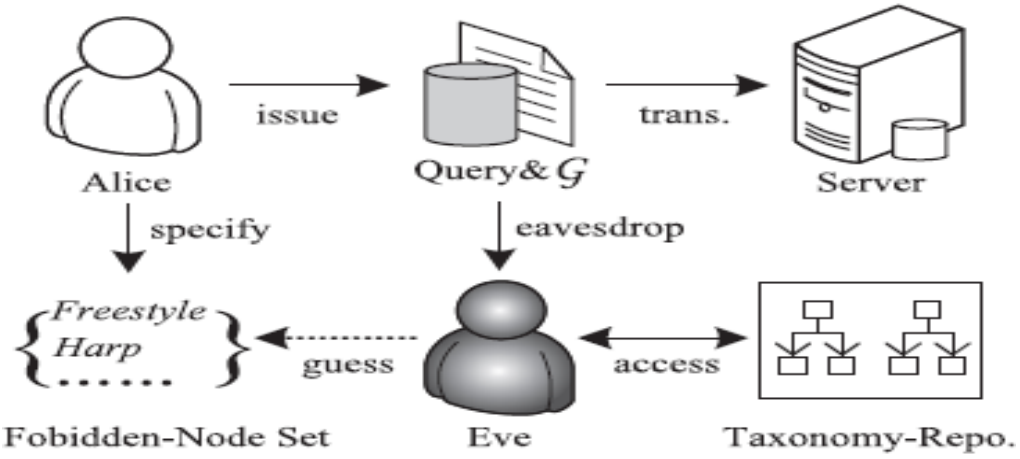

Fig.3. Attack model of personalized web search

Note that in our attack model, Eve is regarded as an adversary satisfying the following assumptions: 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 Session bounded. None of previously captured information is available for tracing the same victim in a long duration. In other words, the eavesdropping will be started and ended within a single query session.

The above assumptions seem strong, but are reasonable in practice. This is due to the fact that the majority of privacy attacks on the web are undertaken by some automatic programs for sending targeted (spam) advertisements to a large amount of PWS-users. These programs rarely act as a real person that collects prolific information of a specific victim for a long time as the latter is much more costly.

5.4 Generalizing User Profile:

Now, we exemplify the inadequacy of forbidding operation. In the sample profile in Fig. 2a, Figure is specified as a sensitive node. Thus, rsbtrðS; HÞ only releases its parent Ice Skating. Unfortunately, an adversary can recover the subtree of Ice Skating relying on the repository shown in Fig. 2b, where Figure is a main branch of Ice Skating besides Speed.

If the probability of touching both branches is equal, the adversary can have 50 percent confidence on Figure. This may lead to high privacy risk if senðFigureÞ is high. A safer solution would remove node Ice Skating in such case for privacy protection. In contrast, it might be unnecessary to remove sensitive nodes with low sensitivity. Therefore, simply forbidding the sensitive topics does not protect the user's privacy needs precisely.

6. CONCLUSION

This paper presented a client-side privacy protection framework called 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 the hierarchical profiles. In addition, UPS also performed online generalization on user profiles to protect the personal privacy without compromising the search quality. We proposed two greedy algorithms, namely GreedyDP and GreedyIL, for the online generalization. Our experimental results revealed that UPS could achieve quality search results while preserving user's customized privacy requirements. The results also confirmed the effectiveness and efficiency of our solution.

REFERENCES

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

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

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

[5] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive Web Search Based on User Profile Constructed without any Effort from Users," Proc. 13th Int'l Conf. World Wide Web (WWW), 2004.