ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

An Efficient Privacy Preserved Personalized Web Search Model using Fully Homomorphic Encryption

Anjali S[1], Reeshma K[2]

Assistant Professor, Dept of ISE, The Oxford College of Engineering, Bangalore, Karnataka, India[1]

Assistant Professor, Dept of ISE, The Oxford College of Engineering, Bangalore, Karnataka, India[2]

ABSTRACT: Personalized web search has shown its effectiveness to "understand exactly what you mean and give you exactly what you want". It is a promising way to improve the accuracy of web search, and has been attracting much attention recently. However, user information are collected and analyzed to fetch the user intention behind the issued query and to provide a better search result. But users' disinclination to disclose their private information during search has become a major challenge for deploying personalized search applications. A PWS framework called User Customizable Privacy Preserving Search (UPS) is used to generalize user profiles by queries with user specified privacy requirements. In order to provide a balance between the utility of personalization and the privacy risk of disclosing the generalized profile this Runtime generalization mechanism is used. Two efficient greedy algorithms, namely Greedy Discriminating power algorithm (GreedyDP) and Greedy Information Loss algorithm (GreedyIL), are used for runtime generalization. In our proposed framework we are enhancing the PWS framework by using homomorphic Encryption of user queries and there by session attacks like eavesdrops attacks are effectively controlled.

KEYWORDS: Personalized web search; Privacy Protection Online profile; Homomorphic Encryption

I. INTRODUCTION

The web search engine has overlong become the most main gateway for common people looking for useful data on the web. It has become very difficult for web search engines to find information that satisfies individual needs of the user as the amount of information on the web continuously increases. As an example different users may use exactly the same query (e.g., "Apple") to search for different information (e.g., the Apple or the Apple Company products), but existing search engines will be giving the same results for these users. (2) A user's information needs may change over time. The same user may use "Apple" sometimes to mean the Apple tab and sometimes to mean the fruit. Existing search engines are unable to distinguish such cases. So without using more user information and/or the search context of a user it is impossible for a search engine to know which sense "Apple" refers to in a query. In order to overcome this problem and to improve the search accuracy, we should utilize more details about user information and personalize search results according to each individual user. So Personalized search is a way to improve search quality by providing search results for people with different information goals with incorporating his/her interests. User information has to be collected and analyzed to find out the motivation behind the issued user query.

PWS systems mainly use click-log-based or profile-based methods to provide better search results. The click-log based methods uses clicked pages in the users query history. But it has strong limitation that it can only work on repeated queries from the same user. Profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. Among the two methods Profile-based PWS has shown more effectiveness by improved web search quality. It will gather personal and behaviour information of users from query history, click-through data, browsing history, bookmarks etc. So clearly, without using more user information and the search context of a user it is impossible for a search engine to know which sense user is seeking a query. Indeed, there

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

is an inherent apprehension between providing personalized search and privacy preservation. Significant gain can be obtained by personalization at the expense of only a small and less-sensitive portion of the user profile known as Generalized profile. However previous works of privacy preserving PWS are far from the optimal.

A Web Search involves interactions between two parties, a user (U) and a search engine (S). There are two basic interactions between a user and a search engine: (1) Search: A user U composes and submits a query q to search engine S, and the search engine S would return some search results $R = \{R_1, ..., R_n\}$ to the user. (2) Browse: A user U chooses to view a result $R_i \in R$, and the search engine would bring the user the content of R_i. So we can tell generally the user identity, queries and viewed result can potentially misused to reveal the user's private life such as political inclination, family life, and hobbies.

II. RELATED WORK

There are several prior attempts on personalizing web search. For Web search applications normally server-client architecture is commonly adopted, where a client (often the web browser) sends queries to a server (the search engine). The search engine analyses the user information need, looks up its index structure of documents, and returns a ranked list of search results to the client for a user to view. Basically there can be server-side personalization, client-side personalization and client-server Cooperative Personalization. Out of these three techniques client-side personalization is more efficient since personally identifiable information is always stored on a user's personal computer. But given a user's query, a client-side personalized search agent can do query expansion to generate a new query before sending the query to the search engine. The personalized search agent can also rerank the search results to match the inferred user preferences after receiving the search results from the search engine. With this architecture, not only the user's search behavior but also his contextual activities (e.g., web pages viewed before) and personal information (e.g., emails, browser bookmarks) could be incorporated into the user profile, allowing for the construction of a much richer user model for personalization. The sensitive contextual information is generally not a major concern since it is strictly stored and used on the client side. Another benefit is that the overhead in computation and storage for personalization can be distributed among the clients.

Many later works on personalized web search focused on how to automatically learn user preferences without any user efforts .User profiles are built in the forms of user interest categories or term lists/vectors. In some earlier studies user profiles were represented by a hierarchical category tree based on ODP and corresponding keywords associated with each category. User profiles were automatically learned from search history. User preferences were built as vectors of distinct terms and constructed by accumulating past preferences, including both long-term and short-term preferences. Privacy concerns are natural and important especially on the Internet. Some prior studies on Private Information Retrieval (PIR), focuses on the problem of allowing the user to retrieve information while keeping the query private. Instead, this study targets preserving privacy of the user profile, while still benefiting from selective access to general information that the user agrees to release.

One class of Privacy protection problem for PWS treats privacy as the identification of an individual. It tries to solve the privacy problem on different levels, pseudo identity, the group identity, no identity, and no personal information. Due to the high cost in communication and cryptography the third and fourth levels are impractical. First level solutions are proved too fragile by generating a group profile of k users and provide online anonymity on user profiles. To shuffle queries among a group of users who issues them useless user profile protocol is also proposed so that entity cannot profile a certain individual. It assumes the existence of a trustworthy third-party anonymizer. Instead of third party to provide a distorted user profile to the search engine legacy social network were used

III. PROBLEM WITH THE EXISTING SYSTEM

The existing profile-based Personalized Web Search systems are not supporting runtime profiling. In these systems a user profile has been generated only once and through offline. Such kind of web search systems is having poor search quality for some ad-hoc queries .So whether to expose the user profile and how much to expose to get the better search result should be considered. Another problem existing is current system do not have the facility of user

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

customization of privacy requirements. This probably makes some user privacy to be overprotected while others insufficiently protected. Also these systems want to interact with the users often while creating the personalized search results. These existing systems normally refine the search results with some metrics such as rank scoring average rank and so on. We need to measure the search quality and breach risk after personalization, with less user interaction. In these systems user sends an online query to the server. The query will be stored in the server and along with that it generates a user profile. Then the server sends the response to the user as per the query. Existing system is not able to effectively protect against the model of privacy attack called eavesdropping. As shown in the figure 1, to attack Alice's privacy, the eavesdropper Eve successfully intercepts the communication between Alice and Server. Whenever Alice will issue a query q to the server the exact query along with the Runtime profile G will be captured by Eve. Based on G Eve can attempt to see the sensitive topics of Alice, whose disclosure introduce privacy risk to the user. In this attack model Eve is assumed as an adversary to be Session Bounded. With this assumption none of previously captured information is available to Eve for tracing the same victim in a long duration. So the eavesdropping will be started and ended within a single query session. This assumption is a strong drawback for these systems in which in all practical situations we cannot guarantee that eavesdropper will be session bounded. The eavesdropper can collect the previous histories of user queries and the user privacy is under risk with user queries only.

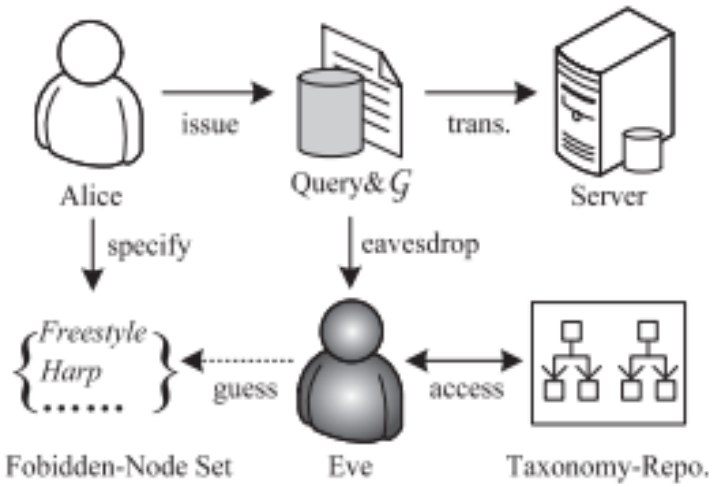

Fig 1. Attack Model

IV. OUR PROPOSED UPS MODEL WITH HOMOMORPHIC ENCRYPTION

All the above problems are addressed in our proposed User customizable Privacy Preserving Search model and aims at protecting the privacy in individual user profile and their queries. As shown in figure 2 the UPS consists of a search engine server and clients. Here online profiler is implemented as a search proxy in the client machine itself. This proxy is used to maintain the complete user profile and customized privacy requirements. It generalizes the profile for each query according to user specified privacy specifications Here a hierarchical user profile is constructed according to user specified privacy requirements. The architectural diagram shown below explains the how the user queries can be handled safely using Homomorphic encryption to avoid. With most encryption schemes the encrypted query has to be decrypted by the server for providing the meaningful search result. This will create extra workload for the server since the computation coast and complexity of the decryption is high. Also and the search result has to be again encrypted and send back to the client. In our model the server computes a succinct encrypted answer without ever looking at the user query in the clear.

- When the user issues a query to the online profiler it generates a user profile in run time based on queries. The generalization process is guided by considering two conflicting metrics, namely the personalization utility and

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

the privacy risk, both defined for user profiles. The output of this is the Generalized User Profile (g) which will satisfy the privacy requirement of the user.

- The original query is passed to the Encryption Module and the user query will be encrypted. The output of this is the Encrypted User query (q').
- The output of the above two modules, Generalized user Profile and Encrypted User Query, will be passed over to the server for personalized web search. The server leave them encrypted, search for them directly in the still-encrypted database with the help of Fully Homomorphic Encryption Mechanism, and get the same results that the client get from the unencrypted data.
- The encrypted search results r' for query are personalized with the generalized profile and delivered back to the proxy.
- Lastly, the proxy either grants the raw results to the user, or re ranks them through the complete user profile after decrypting the search result.

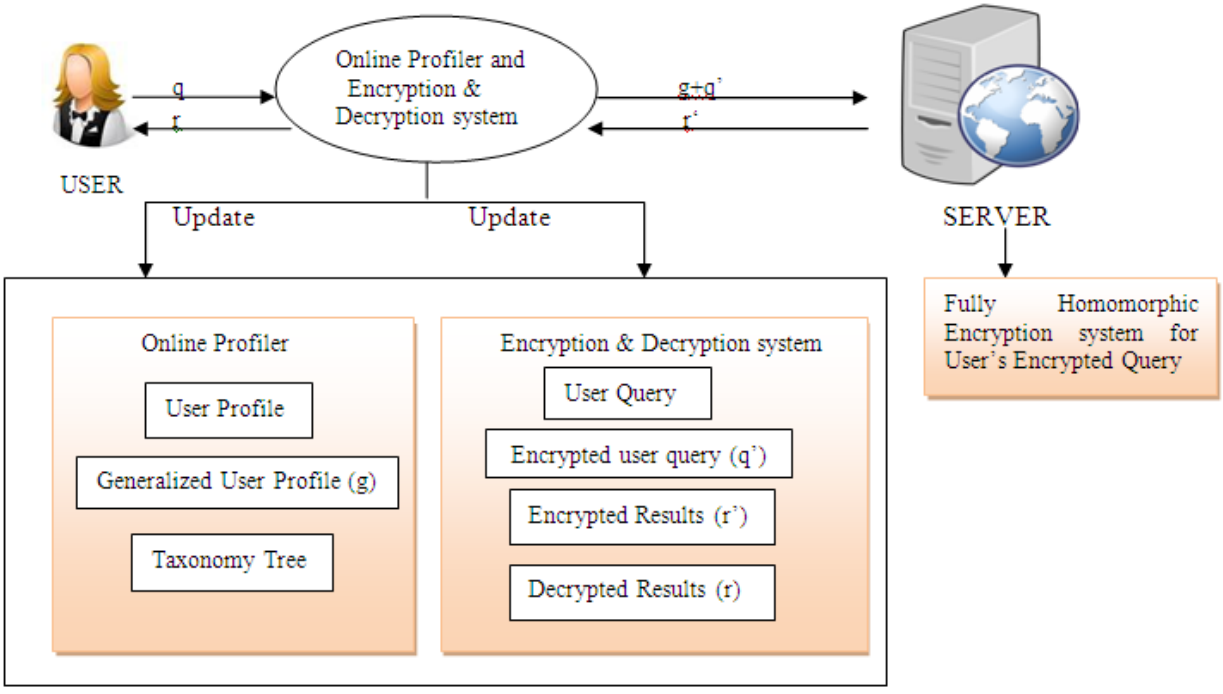

Fig 2. Proposed System Architecture

Fully Homomorphic Encryption (FHE) can be used to query a search engine, without enlightening what is being searched for (here, the search engine is doing the computations on encryptions of information that it doesn't know). More specifically, FHE has the following property in its simplest form. Say that cipher texts c_i decrypt to plaintexts m_i , i.e., Decrypt(c_i) = m_i, where the mi's and c_i's are elements of some ring (with two operations, addition and multiplication). In FHE one has

$$\text{Decrypt}(c_1 + c_2) = m_1 + m_2, \text{Decrypt}(c_1 \cdot c_2) = m_1 \cdot m_2.$$

In other words, decryption is doubly homomorphic, i.e., homomorphic with respect to the two operations addition and multiplication. Being fully homomorphic means that whenever f is a function composed of additions and multiplications in the ring, then

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

Decrypt (f (c_1, . . . , c_t)) = f(m_1, . . . , m_t).

If the adversary can efficiently compute f(c_1, . . . , c_t) from cipher texts c_1, . . . , c_t, without learning any information about the corresponding plaintexts m_1, . . . , m_t, then the system is efficient and secure. Another requirement for FHE is that the cipher text sizes remain bounded, independent of the function f; this is known as the "compact cipher texts" requirement. Fully Homomorphic Encryption schemes can be either public key (where the encryptor knows the decryptor's public key but not her private key) or symmetric key (where the encryptor and decryptor share a key that is used for both encryption and decryption).

V. ALGORITHMS USED

The brute-force algorithm exhausts all possible rooted sub trees of a given user profile to find the optimal generalization. The privacy requirements are respected during the exhaustion. The sub tree with the optimal utility is chosen as the result. Two simple but effective generalization algorithms are used called GreedyDP and GreedyIL. GreedyDP algorithm try to take advantages of Discriminating Power and the second algorithm attempts to minimize the Information Loss (IL). A greedy algorithm is an algorithm that checks the problem resolving heuristic of creating the locally optimum choice at each stage with the confidence of finding a global optimum. GreedyDP works in bottom up manner starting with leaf node, for every iteration, it chooses leaf topic for pruning thus trying to maximize utility of output. During iteration a best profile-so-far is maintained satisfying the Risk constriction. The iteration stops when the root topic is reached. The best profile-so-far is the final result. GreedyIL algorithm improves generalization efficiency. GreedyIL maintains priority queue for candidate prune leaf operator in sliding order. This decreases the computational cost. GreedyIL states to stop the iteration when Risk is satisfied or when there is a single leaf left. Customized Privacy Requirements can be quantified with a number of sensitive nodes (topics) in the user profile, whose revelation (to the server) presents privacy risk to the user.

VI. SIMULATION AND RESULTS

To evaluate the performance of our proposed model, the framework is implemented on a PC with Intel core CPU and 8-GB Main memory. All the mentioned algorithms are implemented in Java language. Tomcat application server is used and Mysql is used as the back end database server. As shown in the figure 3 the performance of GreedyIL is better than the performance of the GreedyDP in terms of response time. GreedyDP requires more time for computing the discriminating power but GreedyIL takes a much smaller response time for generating the results.

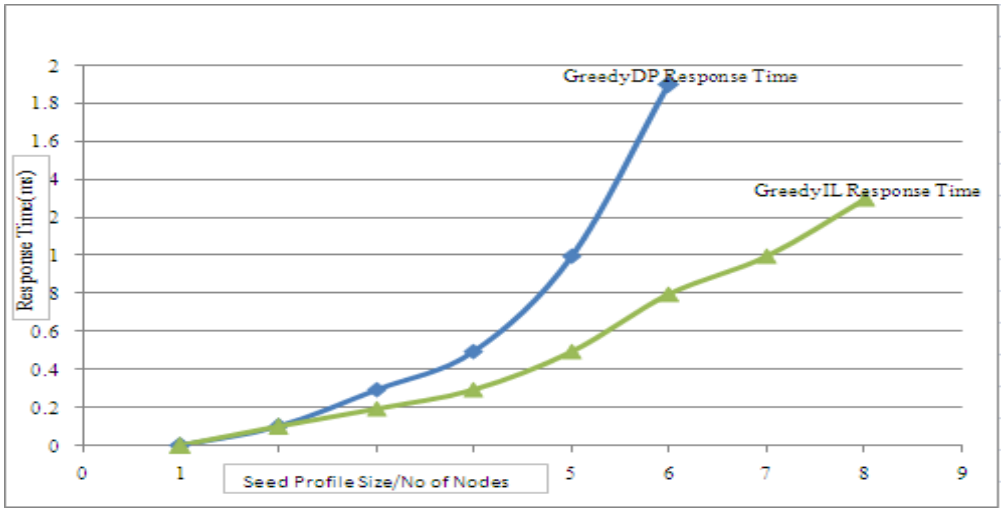

Fig 3. Response time by varying Profile size

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

User customizable Privacy-preserving Search (UPS) framework is created and is used to support privacy in search process as shown in the figure 2. The repository and dataset has been downloaded and kept. The privacy of user query is enhanced with the fully homomorphic encryption scheme. The scalability of the proposed algorithms in terms of response time has been evaluated and the required result has been achieved.

VII. CONCLUSION

The astonishing development of information on the Web has forced new confronts for the construction of effective search engines. This paper presented a new approach to enhance the privacy of the personalized web search systems. The customized privacy requirements of the user are allowed here with effective hierarchical taxonomy. The framework allowed the user to send queries to the search engine in an encrypted manner and avoiding the privacy problems that usually plague online services. The UPS scheme is enhanced with attack resistant from eavesdroppers using Fully Homomorphic Encryption Methods. With this new personalized search engine model we got the personalized search result without revealing any personal information and with an effective attack control rate.

REFERENCES

1. L.Shou,H. Bai,K.Chen, and G. Chen,"Supporting Privacy Protection in Personalized Web Search", IEEE transactions on knowledge an data engineering vol:26 no:2, 2014.
2. M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI),2005.
3. 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.
4. 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.
5. D. Xing, G.-R. Xue, Q. Yang, and Y. Yu, "Deep Classifier: Automatically Categorizing Search Results into Large-Scale Hierarchies," Proc. Int'l Conf. Web Search and Data Mining (WSDM), pp. 139-148, 2008.
6. G.L.Vinay Prasad, Smt S.Jessica Saritha Profile Based UPS personalize Framework for Privacy Protection in Web search. SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE) – volume 2 issue 6, 2015.
7. 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.
8. Ms. P. Sudhaselvanayaki, Dr. T. Senthil Prakash, Ms. V. Karthikeyani, Confidential User Query Profile Construction for Personalized Web Search, International Journal On Engineering Technology and Sciences,2014.

BIOGRAPHY

Anjali S has done her B.Tech in Computer Science and Engineering from College of Engineering Adoor Affiliated to Cochin University of science and Technology and M.Tech in Computer Science and Engineering from AMC College of Engineering Affiliated to Visvesvaraya Technological University. She has 1.7 year industrial experience in company called Isigma Inc. She is currently working as an Assistance Professor in ISE Department of The Oxford College Of Engineering since 1.6 years. She has over all 3 years of experience.

K Reeshma has done her B.Tech in Computer Science and Engineering from College of Engineering Vadakara Affiliated to Cochin University of science and Technology and M.Tech in Computer Science and Engineering from AMC College of Engineering Affiliated to Visvesvaraya Technological University. She has worked at MES Engineering College for three year. She is currently working as the Assistant Professor in ISE Department of The Oxford College Of Engineering since 2 years. She has guided many M.Tech students in Computer Network Engineering. She has over all 5 years of teaching experience.