Asian Journal of Science and Applied Technology (AJSAT)
Vol.3.No.1 2015 pp 34-39
available at: www.goniv.com
Paper Received :08-03-2015
Paper Accepted:20-03-2015
Paper Reviewed by: 1. R. Venkatakrishnan 2. R. Marimuthu
Editor : Prof. P.Muthukumar

CREATING HIERARCHICAL USER PROFILE FOR PRIVACY PROTECTION IN PERSONALIZED WEB SEARCH

[1]J.KAVITHA., [2]S.SUBBIAH M.E.,(PH.D.,),
[1]P.G Student (AU – C), [2]HOD
Department of Computer Science and Engg
Trichy Engineering College
Trichy,India

ABSTRACT

Personalized web search (PWS) is a promising way to improve search quality by customizing search results for people with individual information goals. Personalized web search improves web search by providing content and individual base relation between the search query and its relevant web pages. Here study privacy protection in PWS applications that model user preferences as hierarchical user profiles. Personalized web search is considered a promising solution to improve the performance of generic web search. Now propose a PWS framework called UPS that can adaptively generalize profiles by queries while respecting user specified 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. Here present a greedy algorithms, namely Greedy EHI for optimization. Here enhancing two techniques namely Encoding and Encryption for security. Here use XML files for encrypt the data. So, seek more sophisticated method to build the user profile, and better metrics to predict the performance of UPS.

***Index terms*–Security, privacy risk, profile, personalized web search, optimization.**

I.INTRODUCTION

Personalized web search (PWS) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs. Users might experience failure when search engines return irrelevant results that do not meet their real intentions. Such irrelevance is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of texts. The solutions to PWS can generally be categorized into two types, namely click-log-based methods and profile-based ones. The click-log based methods are straightforward Here simply impose bias to clicked pages in the user's query history. In contrast, profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. Profile-based methods can be potentially effective for almost all sorts of queries. Although there are pros and cons for both types of PWS techniques, the profile-based PWS has demonstrated more effectiveness in improving the quality of web search recently.Unfortunately, such implicitly

collected personal data can easily reveal a gamut of user's private life. To protect user privacy in profile-based PWS, researchers have to consider two contradicting effects during the search process.

On the one hand, attempt to improve the search quality with the personalization utility of the user profile. On the other hand, need to hide the privacy contents existing in the user profile to place the privacy risk under control. In an ideal case, significant gain can be obtained by personalization at the expense of only a small (and less-sensitive) portion of the user profile, namely a generalized profile. Thus, user privacy can be protected without compromising the personalized search quality. Now 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. The existing profile-based Personalized Web Search do not support runtime profiling. A user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminatingly. 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 privacy at risk.

The methods do not take into account the customization of privacy requirements. Many personalization techniques require iterative user interactions when creating personalized search results.so All the sensitive topics are detected using an absolute metric called surprisal based on the information theory. Now 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.

II.EXISTING WORK

In existing model a privacy-preserving personalized web search framework UPS, which can generalize profiles for each query according to user-specified privacy requirements.Relying on the definition of two conflicting metrics, namely personalization utility and privacy risk, for hierarchical user profile, Here formulate the problem of privacy-preserving personalized search as Risk Profile Generalization, with itsNP-hardness proved.Now develop two simple but effective generalization algorithms, GreedyDP and GreedyIL, 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, GreedyIL outperforms GreedyDP significantly.

Now 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.Customized privacy requirements can be specified with a number of sensitive-nodes topics in the user profile,so enhances the stability of the search quality and avoids the unnecessary exposure of the user profile.greedy algorithm GreedyDPnamed as GreedyUtility to support online profiling based on predictive metrics of personalization utility and privacy risk.

III.PROPOSED WORK

Now present a Query Search algorithm, namely Greedy EHI for optimization. Here enhancing two techniques namely Encoding and Encryption for security.

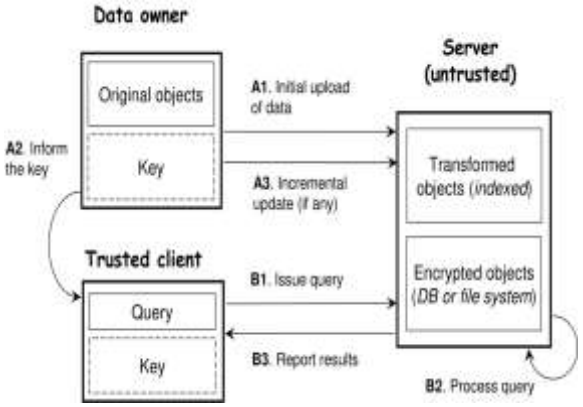

Fig:1 process model

Now use XML files for encrypt the data.so using AES encryption algorithm for encrypt the data and also using two modelanonymization based solution and encryption of the data.

A.OUTSOURCING DATA

It consists of three entities: a data owner, a trusted query user, and an untrusted server. On the one hand, the data owner wishes to upload his data to the server so that users are able to execute queries on those data. On the other hand, the data owner trusts only the users, and nobody else including the server. The data owner has a set P of (original) objects (e.g., actual time series, graphs, strings), and a key to be used for transformation. First, the data owner applies a transformation function (with a key) to convert P into a set P0 of transformed objects, and uploads the set P0 to the server . The server builds an index structure on the set P0 in order to facilitate efficient search. In addition, the data owner applies a standard encryption method on the set of original objects; the resulting encrypted objects (with their IDs) are uploaded to the server and stored in a relational table (or in the file system).

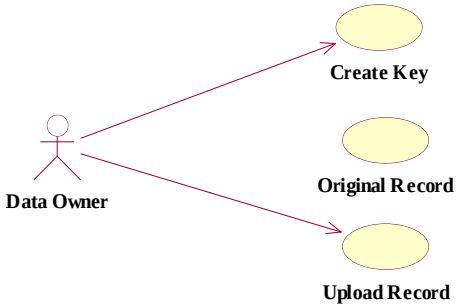

Fig 2: Process of data owner

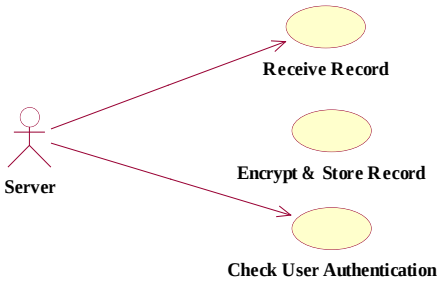

Fig 3: Process of server

B. QUERY UTILITY CREATION

In this module the data owner and the user to apply a transformation function. Our research objective is to design a transformation method that meets the following requirements:

- It enables high query accuracy.
- It enables efficient query processing in terms of communication cost.
- It supports the insertion and deletion of objects.

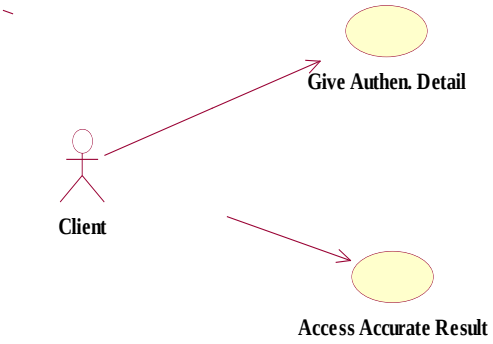

Fig 4: Process of client

C.BRUTE-FORCE SECURE SOLUTION (BRUTE)

This brute-force solution is mentioned in the Introduction. In the offline construction phase, the data owner applies conventional encryption (e.g., AES) on each object and then uploads those encrypted objects to the server. At query time, the user needs to download all encrypted objects from the server, decrypt them and then compute the actual result. As mentioned, it is perfectly secure, but its query and communication cost are both prohibitively high, and pay-as you- go is not supported.

D. ANONYMIZATION BASED SOLUTION

This anonymization-based solution achieves data privacy by means of k-anonymity the objects are generalized in such a way that each generalized object cannot be distinguished from k - 1 other generalized objects. In the context of similarity search, it is able to confuse the ranking of transformed objects because k - 1 of them have the same rank as the transformed object of the actual nearest neighbor. Now still consider this solution as a competitor, even though k-anonymity is not a suitable privacy guarantee for our applications.

E. ENCRYPTION OF THE DATA

This module has to generate encryption task on original data during the data owner store into database. The original data will store in database as encrypted format. It only show the plaintext when the trusted user will access otherwise it will show the cipher text to the un trusted user.Authorized user only know the original data.because permission granted for authorized user decrypt cipertext.then see the original data and upload the website.

F. ALGORITHM QUERY SEARCH ALGORITHM

Algorithm EHI-Search (Query object q, Encryption Key $\mathcal{CK}$, Integer λ)

1: request the server for the (encrypted) root node L_{root};
2: H:=new min-heap; p_{nn}:=NULL;
3: γ:=$\min_{e \in L_{root}} maxdist(q,e)$; ▷ derive NN distance bound
4: **for each** entry $e \in L_{root}$ such that $mindist(q,e) \leq \gamma$ **do**
5: insert the entry $\langle e, mindist(q,e)\rangle$ into H;
6: **while** H is not empty and its top entry's key $\leq \gamma$ **do**
7: pop next λ entries from H and insert them into a set S;
8: request the server for each (encrypted) child node of S;
9: **for each** retrieved node L_{cur} **do**
10: **if** L_{cur} is a leaf node **then** ▷ check for closer objects
11: update γ and p_{nn} by using objects in L_{cur};
12: **else** ▷ expand the entries of L_{cur}
13: γ:=$\min\{\gamma, \min_{e \in L_{cur}} maxdist(q,e)\}$;
14: **for each** $e \in L_{cur}$ such that $mindist(q,e) \leq \gamma$ **do**
15: insert the entry $\langle e, mindist(q,e)\rangle$ into H;
16: **return** p_{nn} as the result;

G.ENCODING OF THE DATA

In this module using XML files for encode the data because the datas stored in the database. This module has to generate encoding task on original data during the data owner store into database. The original data will store in database as 0's and 1's format. It only show the 0's and 1's when the trusted user will access otherwise it will show the 0's and 1's to the un trusted user.Authorized

user only know the original data.because permission granted for authorized user decode the 0's and 1's then see the original data and upload the website.

IV.DATABASE DESIGN

Database design is the process of producing a detailed data model of a database. This logical data model contains all the needed logical and physical design choices and physical storage parameters needed to generate a design in a Data Definition Language, which can then be used to create a database. A fully attributed data model contains detailed attributes for each entity. The term database design can be used to describe many different parts of the design of an overall database system

V.ARCHITECTURE

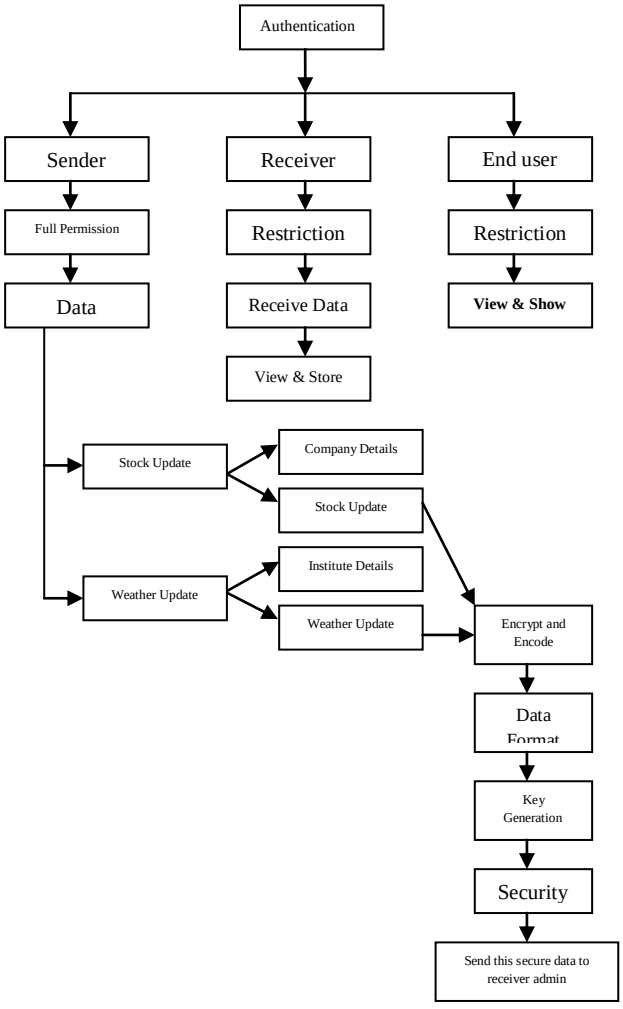

Fig 5: System architecture

VI.SYSTEM DESIGN

Design is multi-step process that focuses on data structure software architecture, procedural details, (algorithms etc.) and interface between modules. The design process also translates the requirements into the presentation of software that can be accessed for quality before coding begins. Computer software design changes continuously as new methods; better analysis and broader understanding evolved.System Design involves the analysis, design, and configuration of the necessary hardware and software components to support your solution's architecture.

The five major components of System Design include: the Information Model, Community Model, Security/Permission Model, System Integration, Workflow, and Technical Architecture.

A.BENEFITS

A System Design engagement typically provides the following benefits:

- Improved system performance; individually tailored configuration advice demonstrates where improvement is necessary, and how to improve the system to regain lost performance.
- Customers gain a detailed understanding of how their users use their system. This Usage Profile can be leveraged to develop future architecture changes.
- Potential to learn of future concerns, allowing customers to take proactive measures to avoid problems.
- A baseline performance level is established against which benefits can be compared and changes to the system predicted or foreseen.

VII.RESULT

The results of the experiment for comparison, Now also plot the theoretical number of iterations of the optimal algorithm. It can be seen that both greedy algorithm outperform Optimal. GreedyDP bounds the search space to the finite-length transitive closure of prune-leaf. GreedyIL further reduces this measure with Heuristic. The greater the privacy threshold, the fewer iterations the algorithm requires. The advantage of GreedyIL over GreedyDP is more obvious in terms of response time.This is because GreedyDP requires much more recomputation of DP, which incurs lots of logarithmic operations. Then Greedy EHI Query Search algorithm is working well and then provide more security using encoding and encryption technique.Then XML files using encrypt the data. So,Here seek more sophisticated method tobuild the user profile, and better metrics to predict theperformance of UPS.

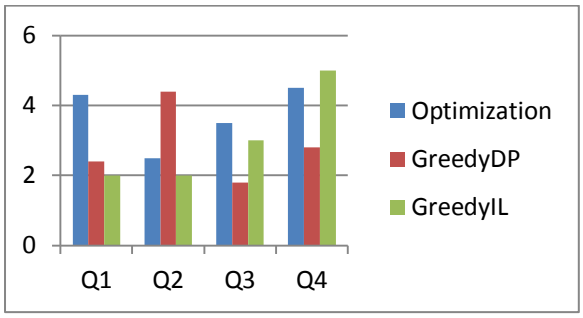

Fig 6:Efficiency of optimization

VIII.CONCLUSION

This paper presented a security and client-side privacy protection for personalized web search. The framework allowed users to specify customized privacy requirements via the hierarchical profiles. Now proposed a greedy algorithms, namely GreedyEHl query Search Algorithm for the optimization. Our experimental result is provide more security and efficiency.Because using encoding and encryption technique for security The results also confirmed the effectiveness and efficiency of our solution.And also using XML files for encrypt the data.So provide hierarchical user profile in Personalized Web Search.

IX.FUTURE ENHANCEMENT

For future work, we will try to resist adversaries with broader background knowledge, such as richer relationship among topics (e.g., exclusiveness, sequentiality, and so on), or capability to capture a series of queries relaxing the second constraint of the adversary from the victim. We will also seek more sophisticated method to build the user profile, and better metrics to predict the performance (especially the utility) of UPS.

REFERENCES

[1] *LidanShou, He Bai, Ke Chen, and GangChen,"SupportingPrivacyProtectioninPers onalizedWebSearch",ieee transactions on knowledge and data engineering, vol. 26, no. 2, february 2014.*

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

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

[4] *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.*

[5] *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.*

[6] *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*

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

[8] *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.*

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

[10] *Y. Xu, K. Wang, B. Zhang, and Z. Chen, "Privacy-Enhancing Personalized Web Search," Proc. 16th Int'l Conf. World Wide Web (WWW), pp. 591-600, 2007.*

[11] *J. Pitkow, H. Schu¨tze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, and T. Breuel, "Personalized Search," Comm. ACM, vol. 45, no. 9, pp. 50-55, 2002.*

[12] *K.Hafner, Researchers Yearn to Use AOL Logs, but They Hesitate, New York Times, Aug. 2006.*

[13] *A.Krause and E. Horvitz, "A Utility-Theoretic Approach to Privacy in Online Services," J. Artificial Intelligence Research, vol. 39, pp. 633-662, 2010.*

[14] *P.A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschu¨tter, "Using ODP Metadata to Personalize Search," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.*

[15] *J.S. Breese, D. Heckerman, and C.M. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Proc. 14th Conf. Uncertainty in Artificial Intelligence (UAI), pp. 43-52, 1998.*

[16] *Pretschner and S. Gauch, "Ontology-Based Personalized Search and Browsing," Proc. IEEE 11th Int'l Conf. Tools with Artificial Intelligence*

(ICTAI '99), 1999.

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

[18]E. Gabrilovich and S. Markovich, "Overcoming the Brittleness Bottleneck Using Wikipedia: Enhancing Text Categorization with Encyclopedic Knowledge," Proc. 21st Nat'l Conf. Artificial Intelli-gence (AAAI), 2006.

[19]J. Castellı´-Roca, A. Viejo, and J. Herrera-Joancomartı´, "Preserving User's Privacy in Web Search Engines," Computer Comm., vol. 32, no. 13/14, pp. 1541-1551, 2009.

[20]Y. Xu, K. Wang, G. Yang, and A.W.-C. Fu, "Online Anonymity for Personalized Web Services," Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 1497-1500, 2009.

[21]K. Ramanathan, J. Giraudi, and A. Gupta, "Creating Hierarchical User Profiles Using Wikipedia," HP Labs, 2008.

[22]R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison Wesley Longman, 1999.

[23]K. Ja¨rvelin and J. Keka¨la¨inen, "IR Evaluation Methods for Retrieving Highly Relevant Documents," Proc. 23rd Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), pp. 41-48, 2000.

[24]X. Shen, B. Tan, and C. Zhai, "Privacy Protection in Personalized Search," SIGIR Forum, vol. 41, no. 1, pp. 4-17, 2007.

[25]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.