ISSN 2395-1621

Enhanced Privacy Protection In Personalised Web Search Using Encryption

[#1]Avan Wansing, [#2]Neha Pandey, [#3]Kamruddin Rogangar,
[#4]Prof. Rajesh Phursule

[1]avanvwansing@gmail.com
[2]nehap7479@gmail.com
[3]kamruddin.rogangar@gmail.com

[#123]Prof. Department of Information Technology
[#4]Prof. Department of Information Technology

JSPM's Imperial College of Engineering & Research, Wagholi, Pune.

ABSTRACT

The retrieval of the data based on user's interest has been demonstrated by Personalized Web Search (PWS) framework. PWS is a general classification of hunt strategies which gives a better list of items which are customized for individual client needs. The user profile is exposed to the server which often violates the user's privacy, So for privacy, personalized web search uses a framework called as User Customizable Privacy Preserving Search (UPS). The generalization algorithms for runtime profiling are GreedyDP which maximize the discriminating power and GreedyIL which minimize the information loss. In this paper, the users profile is encrypted to preserve the privacy of the user. Advanced Encryption Standard (AES) Algorithm for encryption is used. The user will handle the decryption process. User decides which particular information is to be decrypted and which one to not according to his convenience. User information does not get disclosed to the server.

ARTICLE INFO

Article History

Received: 5th May 2016

Received in revised form :

5th May 2016

Accepted: 8th May 2016

Published online :

9th May 2016

Keywords: **Privacy Protection, Personalized Web Search, UPS Framework, Encryption, Generalisation.**

I. INTRODUCTION

These days web crawler has turned into an essential gateway for normal individuals searching for valuable data on the web. On the other hand, clients may encounter disappointment when web indexes return insignificant results that do not meet their genuine aims. Personalized Web Search (PWS) is a general classification of hunt strategies going for giving better list items, which are custom-made for individual client needs. Personalized search refers to a type of search which tailored specifically to an individual's interests by combining information about the individual beyond specific query provided.

One of the solutions to PWS is a profile-based method. In this method, profiles are created for each individual users which help in improving the search quality for various queries. Profile-based Personalized Web Search has a disadvantage that it do not support runtime profiling. A user profile is typically generalized for only once offline and it may not even improve the search quality for some ad hoc queries, exposing user profile to a server has put the user's privacy at risk. This method does not support customization of privacy requirements. This makes some user privacy to be overprotected while others insufficiently protected.

Considering security assurance in PWS applications that model client inclinations as progressive client profiles. For privacy-preserving personalized web search uses a framework called as User Customizable Privacy Preserving Search (UPS).The framework assumes that there's no sensitive information in the queries and aims at protecting the privacy in individual user profiles while retaining their usefulness for PWS. By this profiles are generalize for each query according to user-specified privacy requirements.

Runtime profile goes for striking a harmony between two prescient measurements that assess the utility of personalization and the protection danger of uncovering the summed up profile. The generalization algorithms for Runtime profile are GreedyDP and GreedyIL. First, try to maximize the discriminating power (DP), the latter attempts to minimize the information loss (IL). For security purpose encryption algorithm i.e Advanced Encryption Standard Algorithm is used.

Motivation:

To give user security in profile based PWS, researcher's need to consider two contradicting properties. On the one hand, they attempt to expand the search quality with the

help of user profile while on the other side they have to cover up the protection substance in the user profile .Some of the studies show that the users are willing to compromise privacy for the better search result. In a perfect case, we can have a smooth search result by using a small portion of the user profile, namely a generalized profile. In general, there is a tradeoff between the search quality and level of privacy protection.

Unfortunately, the past work of protection safeguarding PWS is a long way from optimal. The problem with the existing methods is explained in the following observations.

1. The customization of privacy requirement does not consider in the existing system .The user privacy to be overprotected while others inadequately insured. The sensitive topics are recognized utilizing a flat out measurements called surprisal based on information theory, accepting that the less user interest document supports more sensitive. This suspicion can be questioned with a straightforward counterexample: if a client has substantial records about sex the surprisal of this title prompted a conclusion that sex is very general and not sensitive, the fact of the matter is inverse. The prior work can effectively address individual privacy needs during the generalization.

2. While making customized query items many personalization methods require multidrive user interactions. Rank scoring, average rank are generally refined the search result with a few measurements which require multiple user interactions. This worldview is, however, infeasible for runtime profiling as it will not only pose too much risk of privacy breach, additionally request restrictive handling time for profiling. Thus, we need predictive metrics to measure the search quality and breach risk after personalization, without incurring iterative user interaction.

3. In existing system, customized data is not secure from attacker, to solve this problem of privacy data encryption algorithm is used which will encrypt the searched data on user as well as server side.

II. LITERATURE SURVEY

In [1] Personalized Privacy Preservation the author Xiaokui Xiao, Yufei Tao did study on the generalization for preserving the privacy of the sensitive data which is daily produced by the users. The existing techniques concentrate on the each and every approach that cause the same amount of preservation for all the users without analyzing their original needs .This results in providing the insufficient protection to a group of people who actually need it while giving extreme privacy control to the group of people who doesn't need it. This system cannot guarantee the privacy protection in all cases this could lead to cause the unnecessary data loss by performing excessive use of generalization.

Motivating by this limitation, introduced a new generalization framework which takes into account the concept of personalized anonymity. The technique carries out minimum generalization for satisfying users requirements and hold the huge amount of information from the microdata i.e. raw data. However, personalization is an inherent notion of privacy preservation whose objective is to protect the interests of individuals at the first place.

At first, they make a concept that forms a new framework of computing privacy which takes into account the sensible information by an individual preference. Secondly, analyze the theory behind their methodology and evaluate the formulae for quantifying the privacy which clearly show the scenarios where k-anonymity can/can not make sure about safe data production. Finally, they evolved an algorithm for finding the generalized that keeps a huge amount of information in the microdata without breaking any privacy limits. This algorithm optimizes the degree of generalization on quasi-identifier attributes and sensitive attributes.

Advantage-

1. Carry sensible data according to the preference of an individual.
2. Keeps the huge amount of information in the micro data without breaking any privacy limits.
3. It provides privacy protection for the user in almost all cases.
4. There will not be any information loss while performing generalization.

Disadvantage-

1. Provides unnecessary protection to data and generalization.
2. The cost of personalized generalization can be high.

In [2] preserving user's privacy in web search engines the author Jordi Castellà-Roca, Alexandre Viejo, Jordi Herrera-Joancomartí did study on the different web search engine available today on the internet. Web search engines like Yahoo!, Google, Bing, etc. are widely used to find the particular amount of data among a large amount of data in short amount of time. People over the globe use the web search engine for different purposes which are relevant to them. At the same time, needed information belongs to the specific topic is hidden among all the available data and it can be really difficult to find it since that information can be separated all over the World Wide Web. In fact, these useful things can also cause the privacy threats to the users, web search engines can profile the client by storing and analyzing the past queries requested by them. But to solve this privacy threats current mechanism introduces high cost in terms of computation and communication.

In this paper, they produce a novel protocol designed to protect the user's privacy in front of web search profiling. Their system gives the duplicate or deformed user profile to the web search engines. They offered implementation details and computational or communication results which show that the introduced protocol improves the existing solutions in terms of query delay.

The limitation of the existing system was that the person or the entity can get some advantage over the other benefits from the absence of privacy protection mechanism between the user and the web search engine.

Advantage-

1. Provide a duplicate or deformed profile of the user to the search engine.

Disadvantage-

1. The system uses dynamic IP address or proxy which can prevent web search engine to create a profile.

In [3] Privacy-Enhancing Personalized Web Search Benyu Zhang, Zheng Chen Ke Wang, Yabo Xu* has worked on the enhancement of privacy for personalized web search.

Just like the world is growing every second the amount of information on the web also grows, thus satisfies each and every individual users' need by providing the right information is becoming difficult for web search engines. Therefore, search quality can be improved by personalizing the web search quality which is achieved by customizing the search results for people with individual information goals. Users are often uncomfortable in exposing private information to search engines. Therefore, there should be a balance between search quality and privacy protection as privacy can be compromised if there is a gain in service or profit to users.

In this paper, they have presented a scalable way for users to automatically build rich user profiles. According to the users interests, profiles is summarized into a hierarchical organization. The user can choose the content and degree of details of profile information to be exposed to the search engine and in order to achieve this two parameters for providing privacy has been proposed.

Two general approaches are: Search engines returns the query result based on personal information which is re-ranked or sending personal information and queries together to the search engine.

Advantage-

1. Personalized web search improves search quality by customizing the search results for people with individual information goals.
2. This approach finds the feasibility of achieving a balance between users' privacy and search quality.
3. It can work on unstructured data.
4. A balance is achieved between user's privacy requirements and search quality.

Disadvantage-

1. This approach requires users to grant the server full access to personal information on The Internet which violates users' privacy.
2. Cannot find the definition of privacy for unstructured data.

In [4] online offerings, for example, web search, news portals, and e-commerce applications confront the test of giving amazing support of an expansive, heterogeneous client base. To overcome such problem an effort has been introduced by introducing methods to personalize services based on special knowledge about users and their context. Researchers and organizations have sought after explicit and implicit methods for customizing online administrations. For web look, explicit personalization systems depend on clients showing arrangements of subjects of interest that are put away on a server or customer. Implicit systems make utilization of data gathered without client exertion and mindfulness. An approach for explicitly optimizing the utility-privacy tradeoff in personalized services such as web search. Privacy concerns show super-modularity; the more private information we accrue, the faster sensitivity and the risk of identifiability grow.

This methodology depends on two key perceptions. The primary is that for practical applications, the utility picked up with sharing of individual information might frequently have a consistent losses property; gaining more data around a client adds diminishing ads up to the utility of personalization given what is now thought about the client's needs or goals. On the contrary, the more data that is procured around a client, the more concerning of privacy becomes, defining the utility of a set of personal attributes by the focusing power of the information gained with respect to the prediction task. Similarly, the utilization of same probabilistic model to quantify the risk of identifying users given a set of personal attributes. Then combine the utility and cost functions into a single objective function, which is use to find a small set of attributes which maximally improves the probability of anticipating the target website, while making identification of the user as difficult as possible. Overall it is found that significant personalization can be achieved using only a small amount of information about users. Common believe is that the principles and methods employed in the utility-theoretic analysis of tradeoffs for web search have applicability to the personalization of a broad variety of online services.

Advantage-

1. With Super Modular approach more privacy is given to sensitive data of the user.
2. We found that significant personalization can be achieved using only a small amount of information about users.

Disadvantage-

1. The system is dependent on the log of user search activity.

III. PROPOSED SYSTEM

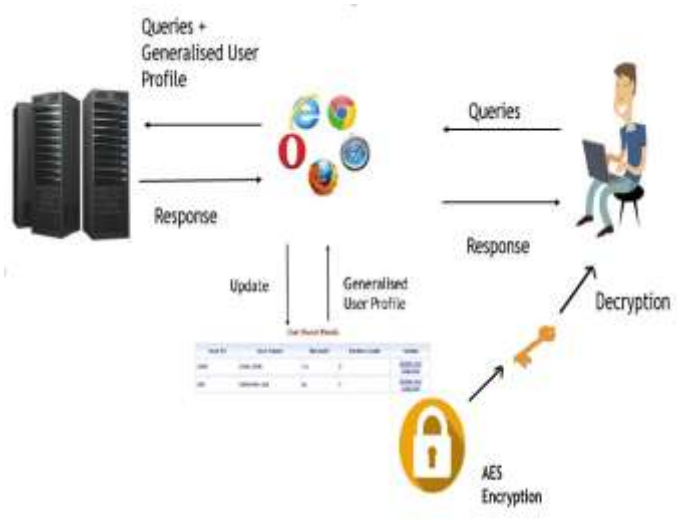

Fig 1. System Architecture

User fires a query on the web. User profile is created at run time, accordingly the user decides which search information is to be encrypted then that search information on the users profile is encrypted by using AES algorithm. Only the user is able to decrypt user profile. The user will get a mail containing One Time Password to decrypt the

information .i.e. the user's privacy is preserved. User information does not get disclosed to the server.

IV. RESULT

Fig 1. Set Privacy Constrain

Fig 2. Search Result

Fig 3. Search Frequency

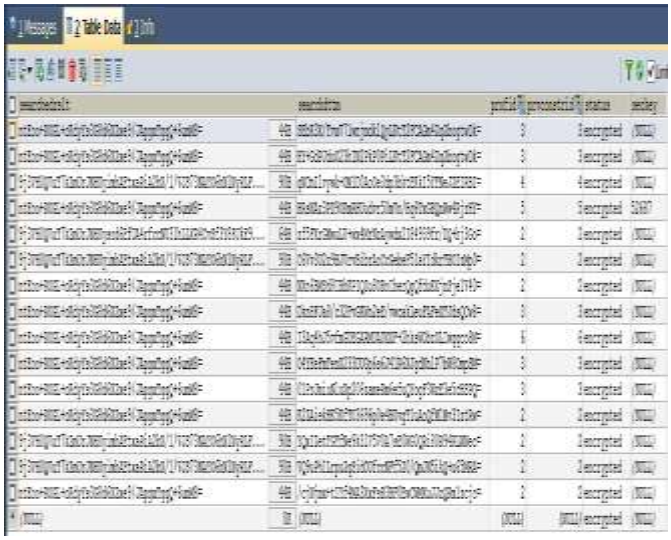

Fig 4. Encrypted Data on Server-side.

V. CONCLUSION

This paper presents 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. The algorithm used for run time profiling is GreedyDP and GreedyIL.

In this paper we have successfully encrypted user profile by using Advance Encryption Standards (AES) Algorithm, which is fast and based on substitution–permutation network. The advance encryption technique was implemented successfully using Java language. This algorithm encrypts the user profile and preserves the privacy of the user. The modification brought about in the code was tested and proof to be accurately encrypting and decrypting the data messages with even high security and immunity against the unauthorized user.

VI. FUTURE WORK

Speed up the process of encryption and decryption. In future work can be done on finding the optimal solution by Dynamic or backtracking method instead of using Greedy method.

REFERENCES

[1] X. Xiao and Y. Tao, "Personalized Privacy Preservation," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2006.

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

[3] 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.

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

[6] Lidan Shou, 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.

[7] Avan Wansing, Neha Pandey and Kamruddin Rogangar "A Survey Paper On Supporting Privacy Protection in Personalized Web Search",IJRET, Vol: 03 Issue:03 Mar-2016.

