International Journal of Engineering and Techniques - Volume 4 Issue 1, Jan – Feb 2018

RESEARCH ARTICLE OPEN ACCESS

Enhancing Personalized Web Search Protection Using Cryptography Algorithm

[1]Mr. S. Dhinakaran, [2]Dr. J. Thirumaran

[1]Research Scholar, Rathinam College of Arts & Science, Coimbatore.

[2]Professor, Dept of Comp. Science,Rathinam College of Arts & Science, Coimbatore.

Abstract:

Over the last twenty years, there has been a extensive growth in the amount of private data collected about individuals. This data comes from a number of sources including medical, financial, library, telephone, and shopping records. Such data can be integrated and analyzed digitally as it's possible due to the rapid growth in database, networking, and computing technologies. On the one hand, this has led to the development of data mining tools that aim to infer useful trends from this data. But, on the other hand, easy access to personal data poses a threat to individual privacy. On the other hand privacy regulations and other privacy concerns may prevent data owners from sharing information for data analysis. In order to share data while preserving privacy data owner must come up with a solution which achieves the dual goal of privacy preservation as well as accurate clustering result.

Some experimental results are presented which tries to finds the optimum value of segment size and quantization parameter which gives optimum in the tradeoff between clustering utility and data privacy in the input dataset. This research work protects the information about PWS applications that model user preferences as hierarchical user profiles.

In this paper, proposes a PWS framework called UPS that can adaptively generalize profiles by queries while respecting user specified privacy requirements. It aims at providing protection against a typical model of privacy attack using the cryptography algorithm.

Keywords — **Privacy, Personalization, Web, Search, Cryptography, Algorithm.**

1. INTRODUCTION

Current web search engines are built to serve all users, independent of the special needs of any individual user. With the exponential growth of the available information on the World Wide Web, a traditional search engine, even if based on sophisticated document indexing algorithms, has difficulty meeting efficiency and effectiveness performance demanded by users searching for relevant information. Personalization of web search is to carry out retrieval for each user incorporating his/her interests. Personalized web search differs from generic web search, which returns identical results to all users for identical queries, regardless of varied user interests and information needs. When queries are issued to search engine, most return the same results to users. In fact, the vast majority of queries to search engines are short and ambiguous. Different users may have completely different information needs and goals when using precisely the same query.

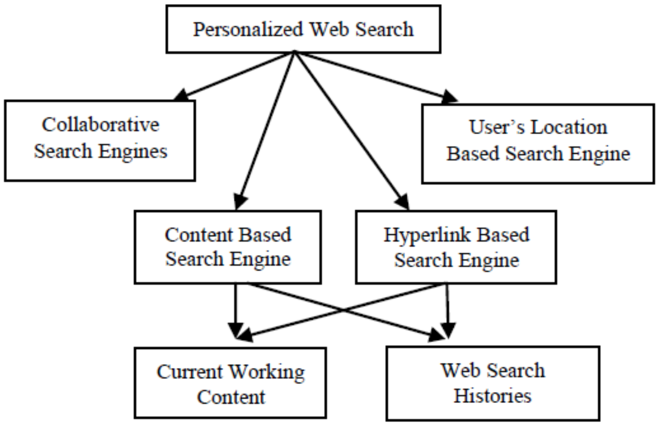
Personalized Web Search
Collaborative Search Engines
User's Location Based Search Engine
Content Based Search Engine
Hyperlink Based Search Engine
Current Working Content
Web Search Histories

Figure 1: - Personalized Web Search Approach

Personalized web search can be achieved by checking content similarity between web pages and user profiles. Some work has represented user interests with topical categories. User's topical interests are either explicitly specified by users themselves, or can be automatically learned by classifying implicit user data. Search results are filtered or re-ranked by checking the similarity of topics between search results and user profiles.

Personalized web search (PWS) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs. As the expense, user information has to be collected and analyzed to figure out the user intention behind the issued query.

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—they simply impose bias to clicked pages in the user's query history. Although this strategy has been demonstrated to perform consistently and considerably well, it can only work on repeated queries from the same user, which is a strong limitation confining its applicability. 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, but are reported to be unstable under some circumstances.

2. BASICS OF PERSONALIZED SEARCH

Generic Search Engines present the results which are general and not adaptable to individual users. For a particular query fired to the search engine, different results are provided for different users. Search results are organized for every user considering one's interest, preferences and information needs. The need for personalization arises due to the following facts: firstly, different users have different backgrounds and interests. For the same query, they have different information needs and goals. Secondly, User information needs may change over time. Users may have variety of requirements based on the time and circumstances.

2.1.1. Creation of User Profile

To provide personalized search results to users, personalized web search maintains a user profile for each individual. A user profile stores information about user interests and preferences. It is generated and updated by exploiting user-related information. Such information may include:

- ➔ Information about the user like age, gender, education, language, country, address, interest areas, and other information.
- ➔ Search history, including previous queries and clicked documents.
- ➔ Other user documents, such as bookmarks, favorite web sites, visited pages, and emails.

2.1.2. Server Side and Client Side Implement

Personalized web search can be implemented on either server side (in the search engine) or client side (in the user's computer or a personalization agent) [1]. For server-side personalization, user profiles are built, updated, and stored on the search engine side. User information is directly incorporated into the ranking process, or is used to help process initial search results. The advantage of this architecture is that the search engine can use all of its resources, for example link structure of the whole web, in its personalization algorithm. Also, the personalization algorithm can be easily adapted without any client efforts. This architecture is adopted by some general search engines such as Google Personalized Search. For client-side personalization, user information is collected and stored on the client side (in the user's computer or a personalization agent), usually by installing a client software or plug-in on a user's computer.

Privacy concerns are also reduced since the user profile 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. A main drawback of personalization on the client side is that the personalization algorithm cannot use some knowledge that is only available on the server side (e.g., PageRank score of a result document). Furthermore, due to the limits of network bandwidth, the client can usually only process limited top results.

2.1.3. Content Based Personalized Search

By checking content similarities between web pages and user profile personalized search can be improved [3]. User's interests can be automatically learned by classifying implicit user data. Search results are filtered or re-ranked by checking the similarity of topics between search results and user profiles. User-issued queries and user-selected documents are categorized into concept hierarchies that are accumulated to generate a user profile. When the user issues a query, each returned result is also classified. The documents are re-ranked based upon how well the document categories match user interest profiles. Chirita et al. [4] use the ODP (Open Directory Project, http://www.dmoz.org/) hierarchy to implement personalized search. User favorite topics nodes are manually specified in the ODP hierarchy. Each document is categorized into one or several topic nodes in the same ODP hierarchy. The distances between the user topic nodes and the document topic nodes are then used to re-rank search results.

2.1.4. Hyperlink Based Personalized Search

Hyperlink Analysis significantly improves the relevance of the web search results so that all major search engines claim to use some type of hyperlink analysis. Web information retrieval mainly focuses on hyperlink structures of the Web, like with Web search engine Google. In personalized Web searches, the hyperlink structures of the Web are also becoming important. The use of personalized PageRank to enable personalized Web searches was first proposed in [6], where it was suggested as a modification of the global PageRank algorithm, which computes a universal notion of importance of a Web page. The computation of (personalized) PageRank scores was not addressed beyond the original algorithm. Experiments [7] concluded that the use of personalized PageRank scores can improve a Web search.

Crawling and ranking are the main uses of hyperlink analysis. In this approach, web crawler which is a software program to browse WWW in automated methodical manner find more and more web pages linked to the source page with the assumption of nearly all the linked web pages are on same topic. This process repeats for each set of web pages until no more linked pages. Then crawler of the search engine orders the web pages by the quality.

In addition to produce a quality and relevant web results, hyperlink analysis have several advantages like finding mirrored hosts, web page categorization and identify the geographical scope of the search etc. But in this approach, search engine has to deal with more details consist even with unnecessary stuffs also. It becomes wastage of the resources.

3. EXISTING APPROACHES & ITS DIFFICULTIES

In the existing approach, propose 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, we formulate the problem of privacy-preserving personalized search as Risk Profile Generalization, with its NP-hardness proved.

It consists of 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. 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.

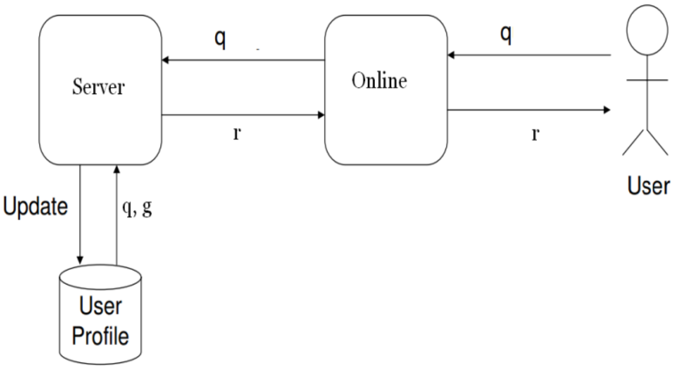

Figure 2: - Generalized approach of Existing System

A diagram of a sample user profile is illustrated in Fig. 2, which is constructed based on the sample taxonomy repository in Fig. 3. We can observe that the owner of this profile is mainly interested in Computer Science and Music, because the major portion of this profile is made up of fragments from taxonomies of these two topics in the sample repository. Some other taxonomy also serves in comprising the profile, for example, Sports and Adults.

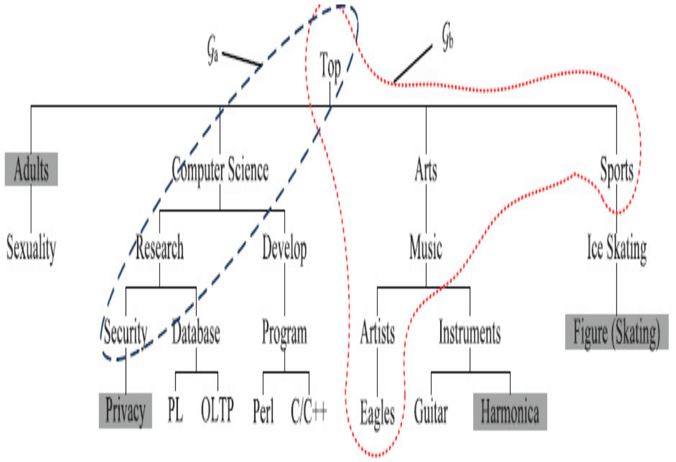

Figure 3: - Sample User Profile

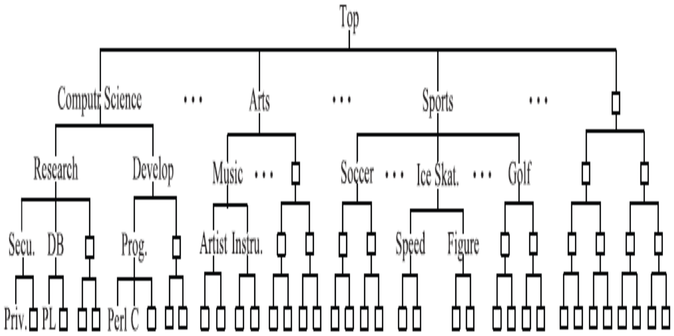

Figure 4: - Sample Taxonomy Repository

The framework works in two phases, namely the offline and online phase, for each user. During the offline phase, a hierarchical user profile is constructed and customized with the user-specified privacy requirements. The online phase handles queries as follows:

1. When a user issues a query qi on the client, the proxy generates a user profile in runtime in the light of query terms. The output of this step is a generalized user profile Gi satisfying the privacy requirements. The generalization process is guided by considering two

conflicting metrics, namely the personalization utility and the privacy risk, both defined for user profiles.

2. Subsequently, the query and the generalized user profile are sent together to the PWS server for personalized search.
3. The search results are personalized with the profile and delivered back to the query proxy.
4. Finally, the proxy either presents the raw results to the user, or re-ranks them with the complete user profile.

The existing scheme can be classified into the following divisions and each division has its own set of processes.

1. Profile Based Personalization
2. Privacy Protection in PWS System
3. Generating User Profile
4. Online Decision

4. PROPOSED SYSTEM AND ITS CONTRIBUTIONS

Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how and to what extent information is communicated to others. Privacy per se is about protecting users' personal information. However, it is users' control that comprises the justification of privacy. With the complete user profile constructed above, an approach without any privacy risk is to grant users full control over the terms in the hierarchy so that they can choose to hide any terms manually as they desire. Unfortunately, studies have shown that the vast majority of users are always reluctant to provide any explicit input on their interests

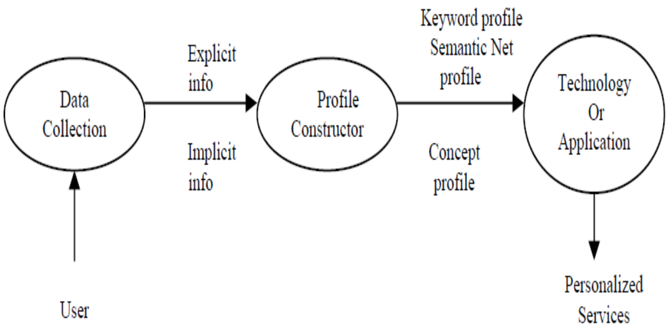

Figure 5: - Overview of user-profile-based personalization

The above Figure 5, the user profiling process generally consists of three main phases. First, an information collection process is used to gather raw information about the user. Depending on the information collection process selected, different types of user data can be extracted. The second phase focuses on user profile construction from the user data. The final phase, in which a technology or application exploits information in the user profile in order to provide personalized services.

4.4.1. Collecting Information about Users

The first phase of a profiling technique collects information about individual users. A basic requirement of such a system is that it must be able to uniquely identify users. The information collected may be explicitly input by the user or implicitly gathered by a software agent. It may be collected on the user's client machine or gathered by the application server itself. Depending on how the information is collected, different data about the users may be extracted. In general, systems that collect implicit information place little or no burden on the user are more likely to be used and, in practice, perform as well or better than those that require specific software to be installed and/or explicit feedback to be collected.

4.4.2. User Profile Construction

User profiles are constructed from information sources using a variety of construction techniques based on machine learning or information retrieval. Depending on the user profile representation desired, different techniques may be appropriate. Profiles may be constructed manually by the users or experts, however, this is difficult and time consuming for most users and would be a barrier to widespread adoption of a personalized service.

4.4.3. Building Concept Profiles

This section describes three representative systems that build user profiles represented as weighted concept hierarchies. Although each uses a different construction methodology, they each use reference taxonomy as the basis of the profile. These profiles differ from semantic network profiles because they describe the profiles in terms of pre-existing concepts, rather than modeling the concepts as part of the user profile itself. Thus, they all require some way of determining which concepts a user is interested in based on their feedback.

4.4.4. ECC Cryptography Algorithm

Elliptic curve cryptography, or ECC is an extension to well-known public key cryptography. In public key cryptography, two keys are used, a public key, which everyone knows, and a private key, which only you know.

To encrypt, the public key is applied to the target information, using a predefined operation (several times), to produce a pseudo-random number. To decrypt, the private key is applied to the pseudo-random number, using a different predefined operation (several times), to get the target information back. The algorithm relies on the fact that encryption is easy, and decryption is hard, making decryption impractical without the key. It was the first system to allow secure information transfer without a shared key.

The problem is that with today's computers getting faster and faster, there will come a point where we can't make the pseudo-prime large enough to thwart an attack. That is where elliptic curve cryptography comes in. This extension uses the properties of an elliptical curve, the same pair of keys, and some funky math (which I won't get into here), to encrypt and decrypt the target information.

The equation of an elliptic curve is given as,

$$y^2 = x^3 + ax + b$$

Few terms that will be used,

E -> Elliptic Curve

P -> Point on the curve

n -> Maximum limit (This should be a prime number).

4.4.4.1. Key Generation

Key generation is an important part where we have to generate both public key and private key. The sender will be encrypting the message with receiver's public key and the receiver will decrypt its private key.

Now, we have to select a number 'd' within the range of 'n'.

Using the following equation we can generate the public key

Q = d * P

d = The random number that we have selected within the range of (1 to n-1). Pis the point on the curve.

'Q' is the public key and 'd' is the private key.

4.4.4.2. Encryption

`Let 'm' be the message that we are sending. We have to represent this message on the curve. This have in-depth implementation details. All the advance research on ECC is done by a company called certicom.

Conside *'m'* has the point *'M'* on the curve *'E'*. Randomly select 'k' from [1 – (n-1)].

Two cipher texts will be generated let it be **C1** and **C2**.

C1 = k*P

C2 = M + k*Q

C1 and C2 will be send.

4.4.4.3. Decryption

We have to get back the message 'm' that was send to us,

M = C2 – d * C1

M is the original message that we have send.

5. CONCLUSION

The remarkable development of information on the Web has forced new challenges for the construction of effective search engines. This research work protects the information on User customizable Privacy preserving Search framework-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.

This research work presented a client-side privacy protection framework called UPS for personalized web search with elliptic curve cryptography algorithm. 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 a public key algorithm called ECC. 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.

6. REFERENCES

[1] G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, "Ups: Efficient Privacy Protection in Personalized Web Search," Proc. 34th Int'l ACM SIGIR Conf. Research and Development in Information, pp. 615- 624, 2011.

[2] 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.

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

[4] Shou, Lidan, et al. "Supporting Privacy Protection in Personalized Web Search."(2012): 1-1.

[5] XuYabo "Online anonymity for personalized web services." Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009.

[6] Shen, Xuehua, Bin Tan, and ChengXiang Zhai. "Privacy protection in personalized search." *ACM SIGIR Forum*. Vol. 41. No. 1. ACM, 2007.

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

[8] 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.

[9] 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.

[10] Sieg, A., B. Mobasher, and R. Burke. "Web search personalization with ontological user profiles", *International Conference on Information and Knowledge Management, Proceedings*, Lisboa, (2007), pp.525-534.

[11] M. Tibouchi. Elligator squared: Uniform points on elliptic curves of prime order as

uniform random strings. IACR Cryptology ePrint Archive, 2014:43, 2014.

[12] K. Rubin and A. Silverberg. Choosing the correct elliptic curve in the cm method. Mathematics of Computation, 79(269):545–561, 2010.

[13] 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.

[14] 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 Year 2014.

[15] 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.

[16] N. Koblitz, “Elliptic curve cryptosystems”, Mathematics of Computation, 48:203-209, 1987.9. Fernandes, F.R.,Machado, R.J.S.,Ferreira, J.M. ,Gericota, M.G. "Gatewaying IEEE 1149.1 and IEEE 1149.7 test access ports "On-Line Testing Symposium (IOLTS), 2012 IEEE 18th International ,June 2012,PP 136 - 137.

[17] G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, “Ups: Efficient Privacy Protection in Personalized Web Search,” Proc. 34th Int‟l ACM SIGIR Conf. Research and Development in Information, pp. 615-624, 2011.