Enable Central Keyword Base Semantic Search Using User Level Encryption Extension Model

K. Muthukumaran, G.S. Rizwana Banu, M.Y. Deepica,
S. Kirithika, P. Manikandan and S. Raju

Abstract--- A cloud system is difficult to synchronize login and authentication data between external clouds and internal systems without exposing internal security data. The cloud technologies are rapidly being adopted throughout the Information Technology (IT) due to their various attractive properties. In spite of their spread, they have raised a range of significant security and privacy concerns which interrupt their adoption in sensitive environments. In existing scheme makes a good trade off between the functionality and the efficiency. To better express the relevance between the query and files, we introduce the TF-IDF rule into our design. Moreover, existing scheme supports both dataset and keywords updates by using the sub-matrix technique. The existing work first gives a basic concept for the central keyword semantic extension ranked scheme, and then proposes two secure searchable encryption schemes to meet different privacy requirements under two different threat models improve. This project presents a comprehensive cloud design which provides an effective and secure cloud computing services on keyword search devices. This proposed system provides fuzzy authorization technique to ensure the correctness of cost and the user level for that fuzzy technique is used. This proposed technique includes if the user level to know the details about the cost that is using the hint given. If the user level is know about the original cost when the re-enter. To avoid this type of the situation user level fuzzy authorization uses the technique if the entered pay cost partially matches the user can get the file search and stop retrying the username and the password. The original data search is taken after TPA verified is given by the real user.

K. Muthukumaran, Assistant Professor, Department of Computer science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal. E-mail:kumarankay@yahoo.com

G.S. Rizwana Banu, Assistant Professor, Department of Computer science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal. E-mail:gsrizwana@gmail.com

M.Y. Deepica, UG Scholar, Department of Computer science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal. E-mail:deepicayuvaraj@gmail.com

S. Kirithika, UG Scholar, Department of Computer science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal. E-mail:skiruthikappm@gmail.com

P. Manikandan, UG Scholar, Department of Computer science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal. E-mail:manipanneer1996@gmail.com

S. Raju, UG Scholar, Department of Computer science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal. E-mail:raju.scse@ksrce.ac.in

I. INTRODUCTION

Cloud Computing

Cloud computing means that instead of all the computer hardware and software you're using sitting on your desktop, or somewhere inside your company's network, it's provided for you as a service by another company and accessed over the Internet, usually in a completely seamless way. Exactly where the hardware and software is located and how it all works doesn't matter to you, the user—it's just somewhere up in the nebulous "cloud" that the Internet represents.

Types of Cloud Computing

1. Infrastructure as a Service (IaaS) means you're buying access to raw computing hardware over the Net, such as servers or storage. Since you buy what you need and pay-as-you-go, this is often referred to as utility computing. Ordinary web hosting is a simple example of IaaS: you pay a monthly subscription or a per-megabyte/gigabyte fee to have a hosting company serve up files for your website from their servers.

2. Software as a Service (SaaS) means you use a complete application running on someone else's system. Web-based email and Google Documents are perhaps the best-known examples. Zoho is another well-known SaaS provider offering a variety of office applications online.
3. Platform as a Service (PaaS) means you develop applications using Web-based tools so they run on systems software and hardware provided by another company. So, for example, you might develop your own ecommerce website but have the whole thing, including the shopping cart, checkout, and payment mechanism running on a merchant's server. Force.com (from salesforce.com) and the Google App Engine are examples of PaaS.

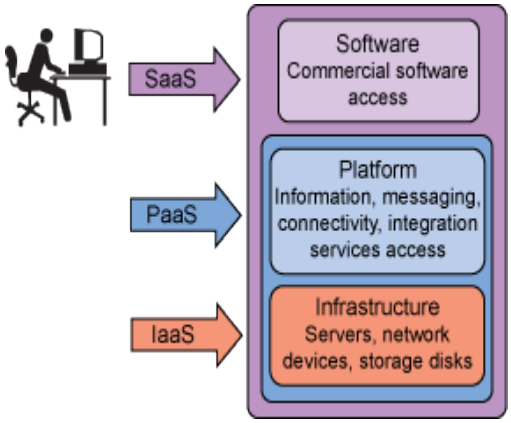

Figure 1.1: Cloud Computing Layers Embedded in the "as a Service" Components

II. LITERATURE SURVEY

Zhicheng Dou, Ruihua Song and Ji-Rong Wen studies this problem and provides some preliminary conclusions. They present a large-scale evaluation framework for personalized search based on query logs, and then evaluate five personalized search strategies (including two click-based and three profile-based ones) using MSN query logs. By analyzing the results, they reveal that personalized search has significant improvement over common web search on some queries but it has little effect on other queries (e.g., queries with small click entropy). It even harms search accuracy under some situations. Furthermore, they show that straight- forward click-based personalization strategies perform consistently and considerably well, while profile-based ones are unstable in their experiments. They also reveal that both long- term and short-term contexts are very important in improving search performance for profile-based personalized search strategies.

1. Personalization may lack effectiveness on some queries, and there is no need for personalization on such queries.
2. Different strategies may have variant effects on different queries. In such a case, simply leveraging pages visited by this user in the past may achieve better performance

III. PURPOSE OF THE PROJECT

The cloud computing becomes more popular nowadays and more users are outsource their data to the cloud, due to its flexibility and unlimited resources. In addition, it can reduce local data maintenance costs and offer a convenient communication channel to share resources among the data owner and legitimate data users.

To enable effective searches over encrypted data, the data owner first builds an encrypted index based on the extracted keywords from data files and the corresponding index-based keyword matching algorithm, and then outsources both the encrypted data and the index structure to the cloud server. To search over the encrypted files, the cloud server integrates the trapdoors of keywords with the index information and finally returns the target files to the data users.

IV. EXISTING SYSTEM

Input central keywords with certain adjunct words as the query keywords when searching documents. The importance of each query keyword depends on the search intension of a user. So far many works have demonstrated the importance of keywords. The super-increasing sequence

is to show the preference factors of keywords to indicate the importance of keywords in a query keyword set.

However, users need to sort keywords according to their importance, which increases the users' input cost. Due to the lack of the super-increasing sequence, the last keyword the user inputs are more important than all the other keywords. The existing model built a user interest model for individual user by analyzing his search history.

However, when inputting unusual keywords, it needs to re-build a new interest model. In this project, our use the grammatical relations as standards to show the weight of each keyword, and this enables users to retrieve relevant documents from the cloud based on their own interests.

Draw Backs of the Existing System

1 Different kinds of access mechanism are not applied and so different client applications with varying processing capabilities need to execute the cloud data in same manner.

2 Session Time limit is not discussed and so client like to access the data in same tariff (cost) for the whole session period.

3 Correlated Authentication aspects with combination of both cloud storage provider, application service provider and end user is not considered.

V. PROPOSED SYSTEM

In addition with all the existing system mechanism, a correlated Authentication aspect with combination of the cloud storage provider, application service provider and end user is also considered. In addition, time limit is provided to end user to access the Application Service Providers (ASPs).Likewise, the security aspects provided by the cloud storage provider is also taken by ASPs to increase the security more. In addition, trusted third party authentication mechanism is included.

Profile Construction

In this module, a hierarchical user profile is constructed and customized with the user-specified privacy requirements.

The first step of the offline processing is to build the original user profile in a topic hierarchy H that reveals user interests. We assume that the user's preferences are represented in a set of plain text documents, denoted by D. To construct the profile, we take the following steps

1. Parse the query.
2. Check the category names which contains query words.
3. Construct the profile tree which contains the category ids mentioned in user/category records along with children

This module handles queries as follows:

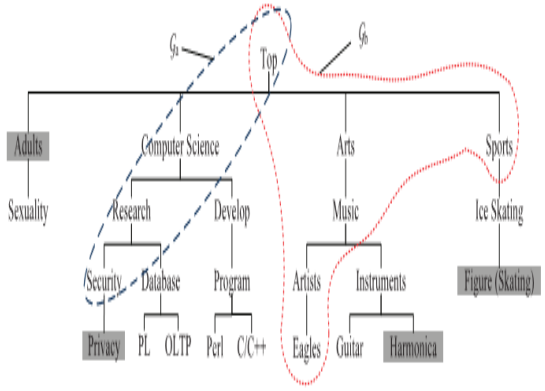

Figure 3.1: Sample User Profile

a. 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 G 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.

b. Subsequently, the query and the generalized user profile are sent together to the PWS server for personalized search.

Search Results

In this module, the search results are personalized with the profile and delivered back to the query proxy.

Re-ranking

Finally, the proxy either presents the raw results to the user, or reranks them with the complete user profile.

Previous Queries

In this module, if the query contains the phrase and the user is already registered (not guest), then the previous queries are also suggested during the query input, so that the search result is related with previous searches.

Advantages of Proposed System

a. Different kinds of access mechanism are applied and so different client applications with varying processing capabilities need to execute the cloud data in same manner.
b. Session Time limit is set and so client likes to access the data in different level for diverse time periods.
c. Trusted third party authentication with no security violation is included.

VI. CONCLUSION

The proposed system is 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. It proposed a greedy algorithm, namely GreedyIL, for the online generalization. The 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. The main benefits are capability to capture a series of queries, User profile is categorized into multiple nodes in the tree structure and past query based suggestion is given to user.

At present, the project presented a client-side privacy protection framework called UPS for personalized web search. For future work, the thesis will try to resist adversaries with broader background knowledge, such as richer relationship among topics (e.g., exclusiveness, sequentially, and so on), or capability to capture a series of queries (relaxing the second constraint of the adversary) from the victim. It 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] 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] R. Krovetz and W.B. Croft, "Lexical ambiguity and information retrieval", Information Systems, Vol.10, No.2, Pp.115–141, 1992.
[3] X. Shen, B. Tan and C. Zhai, "Implicit user modeling for personalized search", Proceedings of CIKM, Pp. 824–831, 2005.
[4] J. Teevan, S.T. Dumais and E. Horvitz, "Beyond the commons: Investigating the value of personalizing web search", Proceedings of PIA, 2005.
[5] X. Shen, B. Tan and C. Zhai, "Context-sensitive information retrieval using implicit feedback", Proceedings of SIGIR, Pp.43–50, 2005.
[6] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories", Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence(WI), 2005.
[7] OnStat, Most people use 2 word phrases in search engines according to onestat.com, 2004. http://www.onstat.com.
[8] Seruku. Seruku toolbar, 2004. http://www.seruku.com/index.html.
[9] Mike Giles, Furl, 2003. http://www.furl.net.
[10] 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.
[11] J. Allan, J. Aslam, N. Belkin, C.Buckley, J. Callan, B. Croft, S. Dumais, N. Fuhr, D. Harman, D.J. Harper and D. Hiemstra, "Challenges in information retrieval and language modeling: report of a workshop held at the center for

intelligent information retrieval", ACM SIGIR Forum, Vol. 37, No. 1, Pp. 31-47, 2002.

[12] T. Joachims, "Optimizing search engines using click through data", Proceedings of SIGKDD, Pp.133–142, 2002.

[13] X. Shen, B. Tan and C. Zhai, "Ucair toolbar: A personalized search toolbar (demo)", Proceedings of SIGIR, 2005.

[14] 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.

[15] R. Cooley, B. Mobasher and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns", Knowledge and Information Systems, Vol. 1, No.1, Pp. 5–32, 1999.