International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Achieving Efficient Multi-Keyword Ranked Search over Encrypted Cloud Data Using Bloom Filters

Sana Shaikh[1], Dr. Rahat Khan[2]

[1]Department of Computer Science and Engineering Marathwada Shikshan Prasark Mandal"s Deogiri Institute of Engineering & Management Studies, Aurangabad Maharashtra state, India 2016-2017

[2]Associate Professor, Department of Computer Science and Engineering Marathwada Shikshan Prasark Mandal"s Deogiri Institute of Engineering & Management Studies, Aurangabad Maharashtra state, India 2016-2017

Abstract: ***In emerging cloud computing, a central application is to outsource the files and record belonging to its user, to outer cloud servers for adaptable information storage. The outsourced documents and files should be encrypted because of the protection and secrecy worries of their proprietor. As there is large amount data present in the cloud it is very important to have a multi-keyword search over the encrypted data. Essentially huge amount of data is present on cloud and providing it for any time on demand request is difficult and is challenging. As searching is time consuming process it is important to provide multi keyword search giving a ranking result to get effective data. To maintain accuracy of search result and also provide better searching experience, it is important for such ranking system to provide multiple keyword searches, as single keyword search gives lots of noisy data. However, for privacy requirement encryption should be done on the sensitive data before outsourcing it, which obsoletes data utilization like information retrieval based on keyword. The main goal of efficient and secure search is building up the searchable encryption for multi-keyword ranked search over the scalable data documents that are stored on cloud.***

Keywords: Cloud computing, Multi-keywords search, Ranking, Privacy requirement, Searchable encryption.

1. Introduction

Cloud computing can be described as a single computing machine or more number of machines which as a server/s, holds all necessary applications and report(information) all at one place so that any user can access it from anywhere whoever has an access to that server without actually using his/her physical machine. There are many cloud platforms present in the market like Google Drive, cloud; SkyDrive, Amazon S3, Dropbox and Microsoft Azure which provide storage services. Security and privacy concerns are the main issues in cloud computing as computing resources are shared many users. Although cloud provider uses hardware and software security mechanisms like firewalls etc. but that is not enough to protect the data. On the other hand, as outsourced information regularly contain touchy security data, for example, individual photographs, messages, and so forth., which would prompt extreme secrecy and protection infringement, if without efficient insurances. It is along these lines important to scramble the touchy information before outsourcing them to the cloud. The information encryption, on the other hand, would bring about notable troubles when different clients need to get to intrigued information with search, because of the challenges of search over encrypted information. This major issue in cloud computing in like manner propels a broad collection of research in the late years on the examination of searchable encryption procedure to accomplish efficient searching over outsourced encrypted information.

2. Literature Survey

In cloud computing, we are outsourcing the data on cloud and when the data is stored in encrypted form, performing the search operation on it is a challenging aspect. In [1] for the first time the matter of secure search over encrypted cloud data is provided by giving a ranked search result. It greatly enhances system usability by providing the files that matches data user"s query in a ranked order regarding to certain relevance criteria (e.g., keyword frequency) thus relevance scoring is done for measuring relevancy using TF-IDF technique, thus going one step closer to practical deployment of secure data hosting services in Cloud Computing. Author gives an ideal construction of ranked single keyword search under the state-of-the-art of searchable symmetric encryption using AES algorithm. The main limitation was it enabled ranked results using single keyword only which gave lots of noisy data too.

In [2] using Searchable symmetric Encryption the data privacy issues are addressed. The scheme provides privacy concern from feature of similarity relevance along with scheme robustness as the author solved the problem of multi keyword retrieving top k result over encrypted data. The sensitive information present on the server-side based on order-preserving encryption which leads on leakage of data. To secure data leakage Two Round Searchable encryption scheme is build. This scheme employs the fully homomorphic encryption also including vector space model, which fulfills the security requirements over encrypted cloud data for multi keyword top-k retrieval. Sufficient search accuracy is provided by vector space model. The homomorphic encryption provides users to engage in the ranking during which the majority of computing work is completed on the server side by operations only on cipher text. The limitation in this paper was security issues such as trapdoor privacy and trapdoor unlink ability was not provided.

In [3] privacy-preserving multi-keyword ranked search supporting similarity-based ranking explained the increasing

popularity of cloud computing, large amount of documents and files are outsourced to the cloud for reducing cost management cost. As encryption helps protecting user data confidentiality, it leaves the well-functioning and secure search functions over encrypted data a challenging problem. In this paper, the principle of "Co-ordinate Matching" is used for search semantic. For privacy preserving multi-keyword scheme the secure k-NN technique is used along with relevance scoring giving similarity-based ranking to address this problem. Author proposed Searchable Symmetric Encryption Scheme to provide the better security using AES algorithm. The paper does not provided data encryption results in difficulties when other users need to access interested data with search as it does not conceal the access pattern of the user.

In [4], using Blind storage for dynamic searchable symmetric encryption, in which client is allowed to store a dynamic collection of encrypted documents with a server, and later perform keyword searches on these encrypted documents, whereas revealing less information to the server. In this paper Author conferred a new dynamic SSE plan that is easy and more efficient than existing schemes while revealing less information to the server than existing system thus enabling fully adaptive security servers. Using Blind storage though provided maximum security but search was not made efficient here. Dynamic searchable symmetric scheme revealing minimum information to cloud server but it fails to give multi keyword search to relinquish efficient ranked result to the user interested query.

In [5], an Efficient as well as Secure Multi-Keyword ranked based Search is proposed. On one side, users who do not essentially have previous information of the encrypted cloud data, have to post process all searched file in order to get ones most matching with the query of their interest; On the other side, consistently retrieving all files containing the query keyword of users interest further pointless network traffic, which is entirely unwanted in today"s pay-as-you-use cloud paradigm. This paper has efficient along with privacy-preserving semantic ranked based search. Latent Semantic search is used where correlation between terms and files are exposed also "k- nearest neighbor (k-NN)" for secure search functionality. Bloom filters are constructed for efficient search. Using Latent semantic Analysis does not only return accurate matching but also will return many relevant files that are semantically co-related. This system fails to give many security issues such as trapdoor privacy, trapdoors unlink ability and concealing access pattern is not considered.

In [6], with the benefits of a public cloud infrastructure, there are also major security and privacy concern. In fact, it seems that the significant obstacle in acceptance of cloud storage is risk over the confidentiality and reliability of data. After the data is encrypted and stored by the data owner this method uses a set of privacy desires for secure cloud data utilization system, by splitting the cloud data and storing the chunk data in different servers. Among different multi-keyword techniques, this method uses the well-organized similarity principle of "coordinate matching" for searching technique. Then the sorted results are created according to top k query scheme. The OTP (One Time Password) is used to observe data in cloud and it can be used only one occasion, when you looking for a file and be inclined to view the file the OTP will transmit to electronic message and you obtain the OTP and be relevant to see the file.

Table 1: Review Summary

Sr.No	Author	Key Techniques	Advantages
1	C. Wang, N. Cao, J. Li, K. Ren, and W. Lou	-Relevance Scoring, -Secure k-nearest neighbour, -Searchable symmetric Encryption (SSE) scheme	This system for the first time gave a secured and ranked keyword search using searchable symmetric encryption scheme.
2	J. Yu, P. Lu, Y. Zhu, G. Xue, and M. Li,	-Relevance Scoring, -Homomorphic encryption scheme, -Vector space model	This system provided multi keyword ranked search accuracy, also enables users to involve in the ranking while the majority of computing work is done on the server.
3	N. Cao, C. Wang, M. Li, K. Ren, and W. Lou,	-Relevance scoring, -secure k-NN search, -Searchable symmetric Encryption scheme. - Co-ordinate matching rule	This paper for the first time provided a privacy preserving multi-keyword ranked search giving trapdoor privacy also.
4	Muhammad Naveed; Manoj Prabhakaran; Carl A. Gunter	-Blind Storage, -Searchable symmetric Encryption Scheme	This paper allows a client to store a dynamic collection of encrypted documents with a server, and later quickly carry out keyword searches on these encrypted documents, while revealing minimal information to the server.
5	Zhihua Xia, Li Chen, Xingming Sun, and Jin Wang	-Latent Semantic Analysis, - k-nearest neighbor (k-NN), -Bloom Filters	This paper provided semantic multi-keyword ranked search scheme for the encrypted cloud data, which simultaneously meets a set of strict privacy requirements.
6	Vanishree R Mr.G.S Suresh	-Relevance Scoring, - Co-ordinate Matching Rule, -One time password(OTP)	This provided multi-keyword ranked search by establishing a set of privacy desires.

3. Existing System Approach

The existing system for information retrieval based on keyword is mainly used on plaintext data, which cannot be applied on encrypted data directly. It is obviously impractical to download the whole data from cloud and decrypt locally. Based on the existence of keywords all these multi keyword search schemes retrieve search result that cannot provide suitable result ranking functionality. However, for privacy requirement encryption should be done on the sensitive data

before outsourcing it, which obsoletes data utilization like information retrieval based on keyword.

4. Key challenges

To achieve efficient and privacy-preserving multi-keyword ranked search over encrypted cloud data via blind storage system, the AMRSB has following design goals:

- *Multi-Keyword Ranked Search*: To meet the necessities for practical uses and endow with better user experience, the AMRSB should support multi-keyword search over encrypted cloud data and achieve relevance-based result ranking.
- *Efficiency of searching*: As the number of the total documents may be huge in a practical situation, the AMRSB should achieve better search efficiency using a sub linear search.
- *Security and Privacy concern*: To put a stop to the cloud server from learning any extra information about the Files and the index, and to maintain search users" trapdoors secret, the AMRSB should cover all the privacy necessities that we introduced above.

5. Privacy Requirement for AMKRSB

- *Files and Index Confidentiality*: Encryption should be performed on both Documents and index before they are being outsourced to a cloud server. The cloud server should be prohibited from snooping into the outsourced files and should not deduce any relationship between the documents and keywords using the index.
- *Trapdoor Privacy*: The Data search user will want her searches to be prevented from being exposed to the cloud server, the cloud server should be not permitted from knowing the keywords enclosed in the trapdoor of the search user.
- *Trapdoor Un-linkability*: The trapdoors must not be linkable; even if they contain the same keywords the trapdoors should be totally different. In other words, rather than being randomized it should be deterministic and should not deduce any links between two trapdoors.
- *Concealing Access Pattern of the Data Search User*: Within multi keyword ranked search, access pattern is the sequence of the searched results of Data Search User. In the AMRSB, the access pattern should be totally covered from the cloud server. In particular, the cloud server should not learn the total number of the documents stored on it nor the size of the searched files even when the Data search user retrieves this file from the cloud server.

6. System Architecture

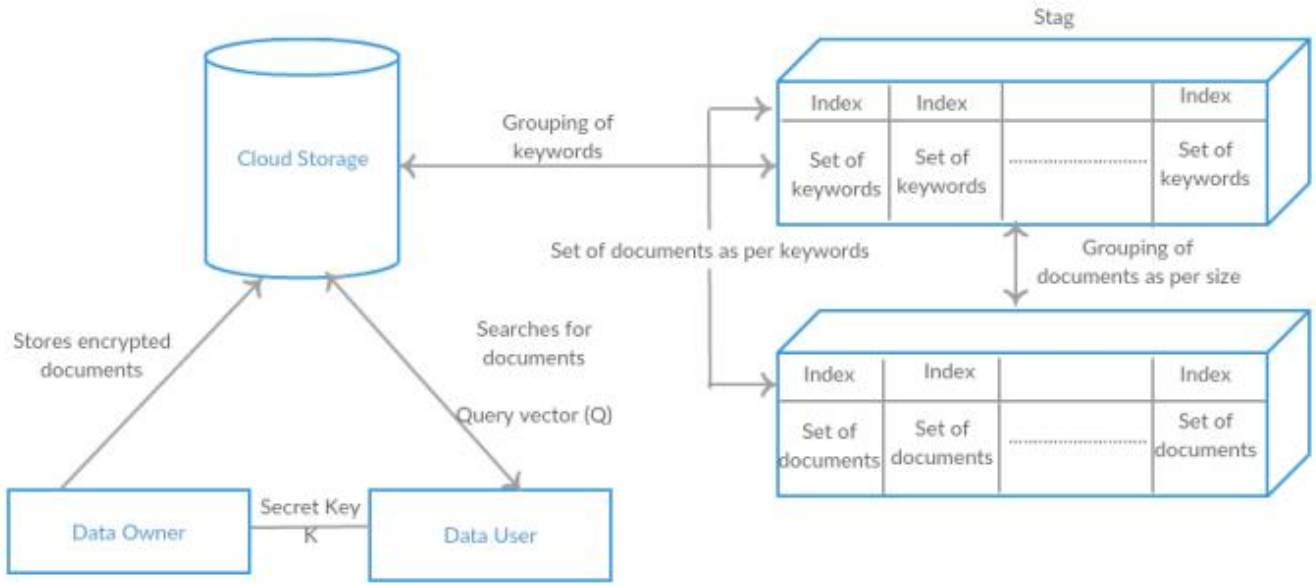

Figure 1: System Architecture

In this system there are two users using the system, namely "The Data Owner" and "The Data User". The Data Owner encrypts and stores data on a cloud server in the form of documents. He also specifies user with whom this files should be share providing them by secret key. Each document that is stored on the cloud along with corresponding index which contains a set of keywords pertaining to each document. This index is also encrypted and stored. As the number of retrieved documents can be large, search results should be retrieved in an order of the relevancy with the searched keywords. Now, when "The Data User" searches for a keyword, the data user should be provided with the number of results that should be displayed as results (the value of k). The search will be done first on the set of encrypted keywords. Before matching the keywords with the searched text, they need to be decrypted. The program then "requests" the system for the secret key which was stored by the "The Data Owner". Upon successful decryption process, the keywords will be matched. All the set of matched keywords will be checked against the corresponding integer and all the relevant top-k documents under that group will be displayed as result.

7. Conclusion

We outline the matter of multi-keyword ranked search over encrypted cloud data. Here we have compared many existing system and algorithm to get a ride over which is more efficient technique to provide a better multi-keyword ranked search with better user experience. However, for privacy requirement encryption should be done on the sensitive data before outsourcing it, which obsoletes data utilization like information retrieval based on keyword.

References

[1] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, „Secure ranked keyword search over encrypted cloud data,“ in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst. (ICDCS). 2010 .pp.253-262

[2] J. Yu, P. Lu, Y. Zhu, G. Xue, and M. Li, „Toward secure multi keyword top-kretrievaloverencryptedclouddata,“IEEETrans.DependableSecure Computer., vol. 10, no. 4, pp. 239–250, Jul./Aug. 2013.

[3] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, „Privacy-preserving multi keyword ranked search over encrypted cloud data,“ IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222–233, Jan. 2014.

[4] Muhammad Naveed; Manoj Prabhakaran; Carl A. Gunter
"Dynamic Searchable Encryption via Blind Storage"
2014 IEEE Symposium on Security and Privacy pp.639-654

[5] Zhihua Xia, Li Chen, Xingming Sun, and Jin Wang "An Efficient and Privacy-Preserving Semantic Multi-Keyword Ranked Search over Encrypted Cloud Data" International Journal of Security and Its Applications Vol.8, No.2 (2014), pp.323-332

[6] Vanishree R , Mr.G.S Suresh "Multi Keyword Ranked Search over Encrypted Cloud Data" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 5, May 2015 pp.2245-2248