International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 06, June-2015

Secure and Integrated Rank Ordered Multiple Keywords Search on Encrypted Data over Cloud

Spurthi P
M.Tech Computer Science & Engg
Bangalore Institute of Technology
Bangalore, India

Dr. Asha. T
Dept.of Computer Science & Engg.
Bangalore Institute of Technology
Bangalore, India

***Abstract*— As cloud computing is having increasing popularity, data owners have been motivated to outsource their sensitive data from local sites to the public cloud for great ease access and reduced cost. To preserve privacy of data, sensitive document has to be encrypted before outsourcing data to cloud. Thus, providing a search facility in an encrypted cloud data is of paramount importance. Since cloud consists of the huge number of data users and documents, it is required to search multiple keywords in the search request and to return documents based on the order of their relevance to these query keywords. The existing paper, for multiple keyword semantics provides a privacy requirement such as "coordinate matching" and "inner product similarity" for securing cloud data. In this paper, we propose integrity checking of rank order for multiple keywords search in an encrypted cloud data. To achieve integrity of rank we use a hash mapping technique so that the order of the rank will be preserved. We establish the strict privacy for securing cloud data utilization system by checking the integrity of rank order.**

Keywords*— *cloud computing, coordinate matching, inner product similarity, integrity of rank.

I. INTRODUCTION

Cloud computing, is an important pattern for highly developed data service, it has become an essential probability for data users to outsource data. It is a service provider based on the demand by the user. Security on cloud includes on outsourcing the sensitive documents such as personal photos, health history and e-mails is increasing hence the sensitive data has to be protected by encrypting them [11]. The cloud service providers are responsible for controlling and monitoring the data, when data owners outsource their sensitive data onto the cloud. Users encrypt their sensitive data before outsourcing it onto cloud, to ensure data privacy, which brings greater challenge for effective data utilization. Cloud customers can store their data remotely to enjoy high quality on-demand services from centralized pool of computing resources [14].

However, even if the cloud storage consists of encrypted data, users have to interact with the cloud so that the cloud can operate on the encrypted data, which causes outflow of sensitive information. Furthermore, in cloud computing, a number of data users, wants to only retrieve the data files in which they are interested in, so that data owners may share their outsourced sensitive data with users. Keyword-based retrieval is one of the most well-liked ways to do so. Keyword-based retrieval it is a representative for data service and it is applied in plaintext, so that users will only retrieve keywords from their interested relevant files from a file set but it is a complicated task in encrypted data scenario due to restricted operations on encrypted data.

In the cloud paradigm, to achieve great flexibility and economic savings, files has to be ranked based on their order of relevance by users interest and the files only with the highest relevancies should be sent back to users and it is required to get the users' interested retrieval result from the relevant files that match instead of all the files [13]. To improve security without compromising with efficiency, we use a scheme such as top-k single keyword retrieval but it does not give the specific document instead it provides relevant many documents which have computation overhead. In top-k multi-keyword scheme over encrypted cloud data each keyword in the search request it helps to narrow down the search result further. But in this paper to refine the result relevance we perform integrity check on the rank order during retrieval of result among multiple keywords over an encrypted cloud data. Multi-keyword semantics retrieval is done by using "Coordinate matching"[12], means as many possible matches in an well-organized similarity measure, and "inner product similarity" means number of query keywords appearing in the document, to quantitatively evaluate the similarity of that document to the search query in "co-ordinate matching" principle. Integrity checking of rank order is done by using hash mapping used to check the integrity of the rank order. Hence checking the integrity of ranked order is important because the order sent by cloud and the order received by data user may be changed by the attack of the intruder.

This paper is organized as follows .In section II, we mention about literature survey. Section III discuss about Problem definition. Section IV is about proposed methodology. Section V is about result and analysis. Section VI discuss about conclusion.

II. LITERATURE SURVEY

Traditional searchable encryption are investigated in[1], [2] and[3] it only supports Boolean keyword retrieval without ranking and also focuses on encryption efficiency and security definitions.

A. Swaminathan, Y. Mao, G.-M. Su, H. Gou, A.L. Varna, S. He, M. Su, and D.W. Oard [4], with improved searchable encryption explores secure rank-ordered retrieval in the situation of the data is centered. For preserving the privacy of

top-k retrieval they built a framework, including ranking with OPE and secure indexing.

As in [6]. They proposed to make indistinguishable relevance scores of different terms they have proposed a function called relevance score transformation and so that indexed data security is improved.

As in [5], they proposed in cloud computing it explored top-k retrieval over encrypted data. By considering the huge number of documents and data users in the cloud, it is required to allow multiple keywords in the search query request and return the documents which are most relevant with these query keywords.

Few existing works as in [7] and [8] they proposed several schemes supporting Boolean multiple keyword retrieval.

In [9], they made their first attempt for defining and solving the problem of top-k retrieval of multiple keywords over encrypted cloud data. To measure and evaluate the relevance score they employ inner product similarity and coordinate matching.

In [10], it explains how to preserve the data privacy so they employed homomorphism. They devised a secure protocol for processing their k-nearest-neighbor index query, so that the query privacy of the client preserving and the data privacy of the owner are preserved.

III. PROBLEM STATEMENT

To design a model that facilitates integrated rank ordered search for efficient utilization of outsourced cloud data, it should simultaneously achieve both security and performance.

- To allow multiple keyword query and provide result with ranking for effective data retrieval.
- Preserving privacy by allowing server to know only about search result and nothing else.
- To achieve low communication and computation overhead.
- To allow the integrated rank order for search results.

IV. PROPOSED METHODOLOGY

The enormous number of on-demand data users and enormous amount of outsourced data documents are present in the cloud, so this problem is challenging as it is really difficult to meet the requirements of scalability, performance and system usability. Ranked search will only send back the most relevant data hence it can also eliminate unnecessary network traffic, in “pay-as-you-use” cloud paradigm which is highly desirable. There are chances of attacking the rank order sent by cloud server hence it is required to check the integrity of rank .To enhance the accuracy of search result as well as to improve the user searching experience, it is necessary to support multiple keywords search with the integrated ranking system.

Fig. 1. Shows the flow of the proposed paper.

Data owner has to perform traditional symmetric key cryptography to encrypt data and then outsource data. Focusing on Index, query and integrated rank ordered it consists of five algorithms.

- Setup – here by taking the security parameters as an input from data owners generating symmetric keys for security is an output to data owners.
- Build Index – here by taking all the documents from the data owners building an index for each and every document by considering unique keywords from the documents. After index construction all the keywords and the documents are encrypted before outsourcing into cloud server.
- Trapdoor – here with the keywords of interest as input, this has to generate trapdoor keys for searching the encrypted cloud data.
- Query – here data user sends the search query to cloud server which consists of multiple keywords and trapdoor. It performs keywords search on the encrypted cloud for the ranked order on the searchable index.
- Integrated Rank – here it checks the integrity of the rank by using hash mapping technique for the integrated ranked ordered search .SHA-1 algorithm is used for checking the integrity of rank. Preserving the integrity of rank is important because when the data user searches for top-k retrieval from the cloud server and when the cloud server returns back the top-k retrieval to the data user there are possibilities of getting attacked by the intruder, so the data user may get the inaccurate rank order. Hence to overcome this problem checking the integrity of rank is required on both cloud and data user. When cloud server sending the top-k retrieval rank order, using rank order it performs a hash mapping technique to generate signature. Then it sends both rank order and signature to the data user. Now the data user by using the rank order generates signature by using same hash mapping technique. Now the data user verifies the signature which is generated by itself and sent by cloud server, if they both matches then the received rank order is accurate else it is inaccurate. If the rank order is accurate then data user downloads the document. If the rank order is inaccurate then he discards the rank order.

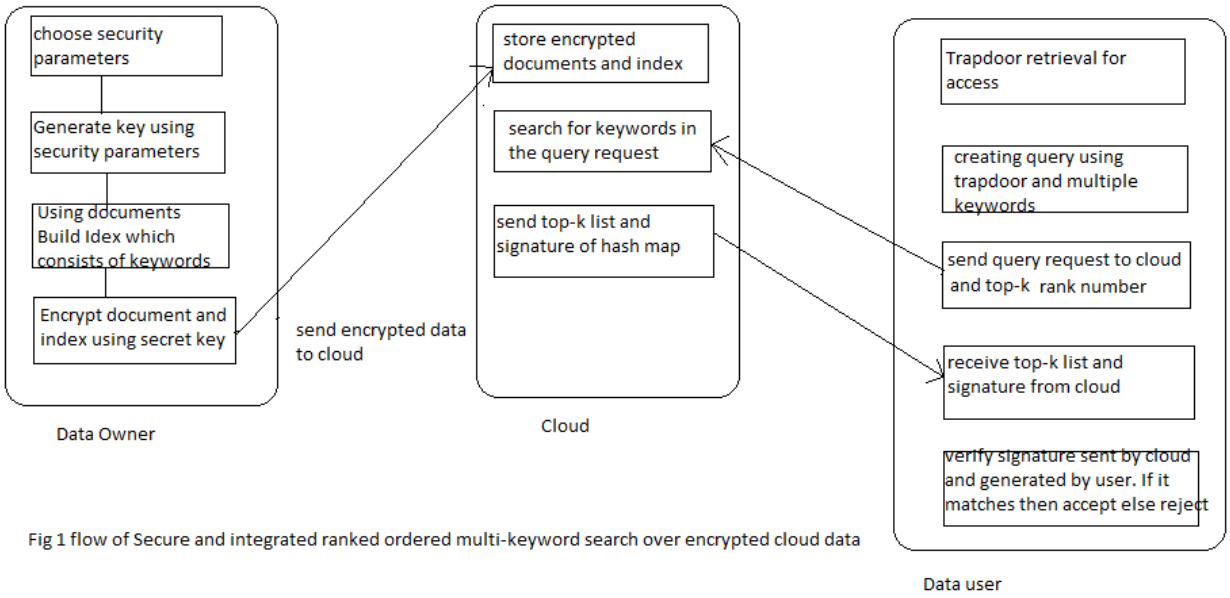
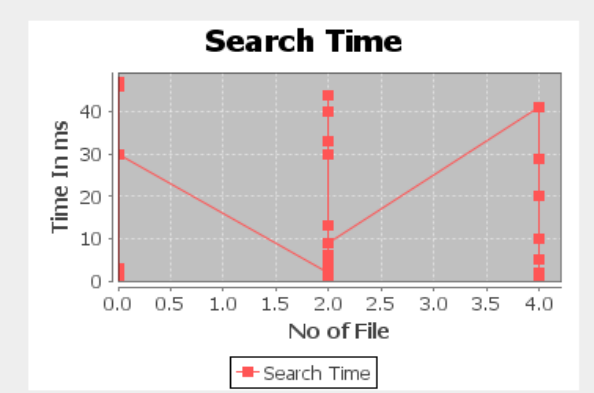


Fig 1 flow of Secure and integrated ranked ordered multi-keyword search over encrypted cloud data

V. RESULT AND ANALYSIS

Fig. 2. graph showing search time result of multi-keyword search without integrity check.



represents keywords present in file

Fig. 2. Multi-keyword search without integrity check

When we search for specific Keywords it searches in all the files and the number of times Keywords appearing in the document will have the highest Rank order.The graph indicates the time taken to search the specific keywords in the document.

Fig. 3. graph showing search time result of multi-keyword search with integrity check.

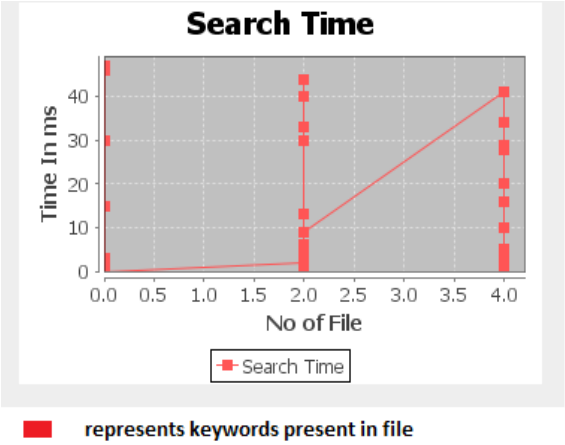


Fig. 3. Multi-keyword search with integrity check

When we search for specific Keywords it searches in all the files and the number of times Keywords appearing in the document will have the highest Rank order. The graph indicates the time taken to search the specific keywords in the document.

VI. CONCLUSION

In this paper we define and solve the problem of checking the security between the cloud server and data user so that the data user retrieves the correct result from the cloud server by using integrated rank ordered multiple keyword searches over encrypted cloud data. We preserve the privacy of the data and the integrity of rank. By using hash mapping technique it solves the challenge of checking the integrity of rank on both cloud server and data user. We can achieve low communication and computation overhead by using an integrated rank order search for effective data retrieval in multiple keyword queries over an encrypted cloud data. As a result of this proposed methodology we can achieve high security while retrieving the data as well as while outsourcing the data.

REFERENCES

[1] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc. ACM 13th Conf. Computer and Comm. Security (CCS), 2006.

[2] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.

[3] D. Boneh, G. Crescenzo, R. Ostrovsky, and G. Persiano, "Public- Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (Eurocrypt), 2004.

[4] A. Swaminathan, Y. Mao, G.-M. Su, H. Gou, A.L. Varna, S. He, M. 5u, and D.W. Oard, "Confidentiality-Preserving Rank-Ordered Search," Proc. Workshop Storage Security and Survivability, 2007.

[5] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int'l Conf. Distributed Computing Systems (ICDCS), 2010.

[6] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k Retrieval from a Confidential Index," Proc. 12th Int'l Conf. Extending Database Technology: Advances in Database Technology (EDBT), 2009.

[7] P. Golle, J. Staddon, and B. Waters, "Secure Conjunctive Keyword Search over Encrypted Data," in Proc. of ACNS, 2004, pp. 31-45.

[8] L. Ballard, S. Kamara, and F. Monrose, "Achieving Efficient Conjunctive Keyword Searches over Encrypted Data," in Proc. of ICICS, 2005.

[9] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multikeyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, 2011.

[10] H. Hu, J. Xu, C. Ren, and B. Choi, "Processing Private Queries over Untrusted Data Cloud through Privacy Homomorphism," Proc. IEEE 27th Int'l Conf. Data Eng. (ICDE), 2011.

[11] S. Kamara and K. Lauter, "Cryptographic cloud storage," in *RLCPS,* January 2010, LNCS. Springer, Heidelberg.

[12] A. Singhal, "Modern information retrieval: A brief overview," *IEEEData Engineering Bulletin*, vol. 24, no. 4, pp. 35–43, 2001.

[13] I. H. Witten, A. Moffat, and T. C. Bell, "Managing gigabytes: Compressingand indexing documents and images," Morgan Kaufmann Publishing,SanFrancisco,May1999.

[14] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," *ACM SIGCOMM Comput.Commun. Rev.*, vol. 39, no. 1, pp. 50–55, 2009.