International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 1 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT1838111

Enhanced Security for Dynamic Multi Keyword Ranked Search Using Greedy Best First Search

Bibin Baby[1], Sharmila Banu[2]
[1]Assistant Professor, Department of Computer Science, Nilgiri College of Arts and Science, Thaloor, The Nilgris, Tamil Nadu, India
[2]Assistant Professor, Department of Computer Science, Providence College for Women, Coonor, The Nilgiris, Tamil Nadu, India

ABSTRACT

Today, due to the enormous growth of data technology in cloud computing, the data owners are stimulated to outsource their data in data management to reduce cost and for the convenient. Data confidentiality, in general, can be obtained by encrypting the data before it is outsourced. The client stores the encrypted data to the cloud using Searchable encryption schemes and applies keyword search techniques over cipher text domain. But the main problem in outsourcing is the lack of security and privacy for the sensitive data. So, to overcome this, for privacy requirement, the sensitive data can be encrypted before it is outsourced. Various methods were proposed to preserve the privacy and to provide security to the cloud data which are encrypted. Here in this paper, we proposed a tree-based search method over the encrypted datain the cloud that supports dynamic operation and multi-keyword ranked search. Clearly, the commonly used "inverse document frequency (IDF) term frequency (TF)" model and the vector space method are joined in the query generation and index creation to give multi-keyword ranked search. To get high search efficiency, a tree-type index structure, "Greedy Best-first Search" algorithm is proposed based on the tree- index.

Keywords : Cloud Computing, Security, Multi Keyword Search, Greedy Best First Search, Inverse Document Frequency, Term Frequency

I. INTRODUCTION

Nowadays, cloud computing has been taken into consideration as a new edition of IT infrastructure, that may put together large beneficial useful resource of computing, storage and programs, and allow users to revel in ubiquitous, convenient and on-demand network get admission to a shared pool of configurable computing sources with splendid overall performance and minimal monetary overhead .Attracted by way of those appealing features, each individuals and firms are stimulated to outsource their facts to the cloud, rather of buying software program and hardware to manipulate the facts themselves. No matter the various blessings of cloud services, outsourcing touchy information (which includes e-mails, non-public health statistics, enterprise business enterprise finance records, authority's files, and plenty of others) to far flung servers brings privacy worries. The Cloud Carrier Vendors (CSVS) that preserve the facts for users may also additionally get admission to users' sensitive information without authorization. A fashionable technique to shield the information confidentiality is to encrypt the statistics earlier than outsourcing but, this will motive a big rate in phrases of records usability. For an example, the present techniques on keyword-based totally information retrieval, which are widely used on the plaintext records, cannot be immediately implemented at the encrypted records. Downloading all of the statistics from the cloud and decrypt regionally is impractical. On the opposite, greater sensible unique motive solutions, including searchable encryption (SE) schemes have made unique contributions in phrases of performance, capability and safety.

II. RELATED WORKS

This section discusses the literature survey of various papers related to Multi-keyword ranked search over encrypted cloud data in cloud. It describes the research work already done by the authors corresponding to searchable encryption in cloud computing.

Qin Liu [3] proposed the pursuit that gives catch phrase protection, information protection and semantic secure by open key encryption. The main drawback of this system is the correspondence and computational expense for encryption and decoding is more. Cong Wang [4] proposed seek which unravels preparing overhead, information and catchphrase security, least correspondence and calculation aeronautical. The information proprietor assemble list alongside the catchphrase recurrence based importance scores for documents. Butit doesn't play out numerous catch phrase seeks and Minimal overhead in record building. C Wang et al [5] proposed that searchable encryption idea agree to the client to solidly search for over scrambled information through watchwords without first applying decoding on it. But this system is too slow in processing the request and is not suitable for huge volume of data.

Wenhai Sun [6] proposed an inquiry that gives comparability based item positioning, catchphrase security, Index and Query classification and Query Unlink capacity. The scrambled record is worked by vector space model supporting solidified and particular document look. The searchable file is fabricated utilizing Multidimensional B tree. Its limitations are the likeness rank score of the record vector completely relies on upon the kind of the report Jiadi [7] proposed this pursuit utilizing two round searchable encryption (TRSE). The withdrawal and restricting is utilized to diminish figure content size, still the key size is too extensive. The correspondence elevated will be high, if the scrambled trapdoor's size is too vast. It doesn't make powerful searchable file redesign. These are some of the drawbacks of this method.

III. PROPOSED METHODOLOGY

In this section, the Dynamic Multikeyword Ranked Search for Plaintext (DMRSP)method is explained for implementing our proposed method. Based on the vector space model and Best First tree based indexing, DMRSP is constructed. Apart from this, a secure search scheme, Simple DMRS (SDMRS) is also explained.

3.1 Index Construction using Best First Search

.

The search algorithm, **Best-first search** explores a graph by increasing the most likely node chosen as per the specified rule.

Judea Pearl explains best-first search as calculating the promise of node '*n*' by a "heuristic evaluation function. It may depend on the depiction of *n*, the goal and the information gathered. It also collects additional knowledge about the problem domain.

A priority queue is the efficient way of selecting the best candidate.

The first successor of the parent is expanded using greedy algorithm. Once a successor is generated,

1. If the heuristic of successor is best, then the successor is put at the front of the queue and restarts the loops.
2. or else, insert the successor into the queue

The algorithm for Best first search is explained below for the following fig. 3.1

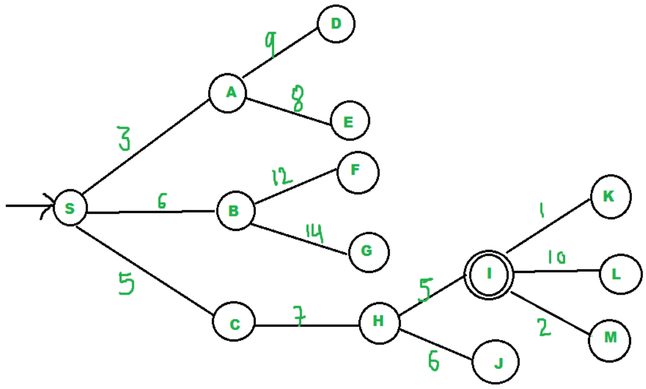

Fig. 3.1 Best First Search Tree

1. We start from source S and search for destination Ith node using Best First search with the given cost.
2. P and Q initially contains S
3. Remove S and process unvisited neighbors of S to pq
4. Now pq contains {A, C, B} (C is put before B because C has lesser cost)
5. Remove A from pq and process unvisited neighbors of A to pq.
6. Now pq now contains {C, B, E, D}
7. Remove C from pq and process unvisited neighbors of C to pq.
8. Now pq contains {B, H, E, D}
9. Remove B from pq and process unvisited neighbors of B to pq.
10. Now pq contains {H, E, D, F, G}
11. Remove H from pq.
12 Now return I

Worst case time complexity : Best First Search : O(n * Log n)

- n is number of nodes.

For achieving this, we need to visit all nodes before we reach the destination. Since priority queue uses Max or Min heap, insert and delete operations takes O(log n).So the performance of the algorithm highly depends on evaluation function or the cost is designed.

IV. RESULT DISCUSSION

These modules are implemented using Java and MySQL. For accessing the location, we have implemented using J2ME as a mobile application. The experiments are carried out on Intel Core(TM) I3 Processor (2.27 GHz).

4.1 Precision

In the proposed method, dummy keywords affects the search precision. Search Precision can be defined as

Pkd=kd'= k wherekd' = top k- documents

To obtain higher precision, the very small standard deviation σ is set as the random variable for $\sum \varepsilon_v$

Table 4.1 Storage Consumption of Index Tree

Size of the dictionary	**1000**	**2000**	**3000**	**4000**	**5000**
GDFS (MB)	78	151	230	335	400
BFS (MB)	72	137	210	297	380

From the above table it is clearly noted that, the memory occupied by Best first search is very less compared to the greedy depth first search.

Table 4.2 Precision Percentages

NO	Precision	NO	Precision
1	89%	9	90%
2	95%	10	96%
3	96%	11	96%
4	97%	12	87%
5	100%	13	100%
6	86%	14	100%
7	88%	15	95%
8	90%	16	87%

We have randomly chosen 5 keywords as input for each test and the precision is noted for the top 100 results. This test is repeated 16 times and the average precision obtained is 93%. The below table gives the precision percentage obtained for multi-keyword search on searching the documents.

V. CONCLUSION AND FUTURE WORK

A tree-based search method is proposed over the encrypted data in the cloud that supports dynamic operation and multi-keyword ranked search on the document collection. Precisely, the widely-used “term frequency (TF) ×inverse document frequency (IDF)” model and the vector space model are combined in the index creation and query generation to afford multi-keyword ranked search. In order to get high search efficiency, a tree-based index structure is constructed and for that a “Best First Search” algorithm is proposed based on this tree index. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. From the results, it is clearly states that the special structure, tree-based index can easily achieve sub-linear search time and compact with the insertion and deletion of documents. Due to the tree-based index, the search complexity of the proposed work is basically kept to logarithmic. The proposed scheme, in practice, can achieve greater search efficiency by implementing our “Best first Search” algorithm.Even though our proposed method works efficiently in terms of multi-keyword search, still there are some serious issues which we will do in the future work. For instance, every time data owner generates updating information and this will be sent to the cloud server. So he has to store the plaintext tree index and IDF values for recalculation.

VI. REFERENCES

[1]. Q. Liu, G. Wang, and J. Wu, "Secure and privacy preserving keyword searching for cloud storage services,' Journal of Networks and Computer Applications, vol. 35, no. 3, pp. 927–933, May 2012

[2]. C. Wang, N. Cao, K. Ren and W. Lou, -Enabling secure and efficient ranked keyword search over outsourced cloud data, IEEE Transactions on Parallel and Distributed Systems, vol. 23, issue 8, (2012) August, pp. 1467–1479.

[3]. C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data,' in Proc. of ICDCS'10, 2010.

[4]. W. Sun, B. Wang, N. Cao, M. Li, W. Lou, YT, Hou and H. L, -Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking, In Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security, (2013), pp. 71-82.

[5]. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data,' in IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, 2014.

[6]. Yu, Jiadi& Lu, Peng& Zhu, Yanmin&Xue, Guangtao& Li, Minglu. (2013). Towards Secure Multi-Keyword Top-k Retrieval over Encrypted Cloud Data. IEEE Transactions on

Dependable and Secure Computing. 10.1109/TDSC.2013.9.

[7]. W. Sun, B. Wang, N. Cao, M. Li, W. Lou, YT ,Hou and H. L, -Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking, In Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security, (2013), pp. 71-82.

[8]. Cengizorencik , ErkaySavaş, An efficient privacy-preserving multi-keyword search over encrypted cloud data with ranking, Distributed and Parallel Databases, v.32 n.1, p.119-160, March 2014

[9]. Chen, Lanxiang, et al. "DMRS: an efficient dynamic multi-keyword ranked search over encrypted cloud data." Soft Computing 21.16 (2017): 4829-4841

[10]. C Orencik, M Kantarcioglu, E SavasA practical and secure multi-keyword search method over encrypted cloud data 2013 IEEE Sixth International Conference on Cloud Computing, 390-397

[11]. Peng, Tao, Qin Liu, and Guojun Wang. "a multilevel access Control scheme for Data security in Transparent Computing." Computing in Science & Engineering (2016).

[12]. Shen, Jian, et al. "An authorized identity authentication-based data access control scheme in cloud." Advanced Communication Technology (ICACT), 2016 18th International Conference on. IEEE, 2016.

[13]. Shen, Zhirong, JiwuShu, and Wei Xue. "Keyword Search With Access Control Over Encrypted Cloud Data." IEEE Sensors Journal 17.3 (2017): 858-868

Cite this article as :

Bibin Baby, Sharmila Banu, "Enhanced Security for Dynamic Multi Keyword Ranked Search Using Greedy Best First Search", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 1, pp. 74-78, January-February 2019.

Available at doi :
https://doi.org/10.32628/CSEIT1838111
Journal URL : http://ijsrcseit.com/CSEIT1838111