Enabling Fine Grained Multi-Keyword Similarity Search Over Encrypted Cloud Data

[1]K.Sujith Kumar,M.Archana[2]

[1]PG Scholar, Dept of CSE, CVR College Of Engineering, Vastunagar, Mangalpalli,Ibrahimpatnam, Ranga Reddy, Telangana, India.

[2]Assistant Professor, Dept of CSE, CVR College Of Engineering, Vastunagar,Mangalpalli, Ibrahimpatnam, Ranga Reddy, Telangana, India.

[1]dukesujith743@gmail.com,[2]mogullaarchana23@gmail.com

***Abstract*:** Cloud computing has brought a drastic change in all aspects especially interms of cost effective for its great flexibility. Considering large number of information users and data owner's searchable encryptions are carried out based on symmetric key on single keyword. In this work multiple keywords to challenge privacy preserving. For multi-keyword search of indexing Bloom Filter is implemented, for privacy concern of query and for fast retrieval, to introduce a special tree-based index structure using random traversal algorithm, which searches easiest path in different way for paths on the index, and guarantee the accuracy under a strict stronger privacy. Finallyit combines these methods to ensure strong approach in security and efficiency to address Bloom filter.

Keywords: cloud computing, data encryption, access control, symmetric key.

1. INTRODUCTION

Cloud computing gives us methods by whichone can get applications as utilities, over the Ib. It enables us to make, design, and search applications on the Ib. Asset partaking in an unadulterated attachment and play display that drastically streamlines foundation arranging is the guarantee of cloud computing. With the approach of this innovation, the cost of calculation, application facilitating, content stockpiling and conveyance is lessened altogether. Distributed computing is a reasonable way to deal with encounter coordinate money saving advantages and it can possibly change a server farm from a capital-concentrated set up to a variable estimated condition. Cloud processing depends on an extremely key central of reusability of IT abilities. But individually, they also need to concern of outsourced data because they contain sensitive information which was used by their applications. Since the investigation of these informational indexes may give significant bits of knowledge into various key zones in the public eye consequently information proprietors require successful, versatile and security safeguarding administrations before discharging their information to the cloud. It is very difficult to retrieve Data from encrypted files. Especially in storing cloud data, this is because of rapid increase in demand, users as their files. Thus it is very complex to reach the requirements of both system performance and retrieval feasibility.

In modern era, majority of the information is transmitted in text only. In this Information retrieval keyword plays an important role. If the Data user wants any file they searches with one or more keywords, if the exact match is not found then it searches for the relevant or similar word. Hence there can be of large number of applications with the similar keyword which results in record linkage21 and biological database32. As this search is conducted for many times but only a few are done under constraint of cipher text all remaining are without security and privacy concerns.

2. EXISTING SYSTEM

K.Ren et.al[6]cloud computing claims to the present most energizing registering change in outlook in innovation of the data. Be that as it may, security and protection are as essential snags to its wide selection. Here, the creator's layouts a few basic securities challenges and inspire facilitate examination of security answers for a reliable open cloud condition.

S. Kamara and K. Lauter [8] consider the issue of building a safe cloud storage benefit over an open cloud in-frastructure where the specialist organization isn't totally trusted by the client. In this work at an abnormal state, a few designs that join later and non-standard cryptographic tidy actives by keeping in mind the end goal to accomplish our objective. we study the benefits such as an engineering would give to the two clients and specialist co-ops and give an outline of late advances in cryptography inspired specically by cloud storage.

A. Broder and M. Mitzenmacher,[9] A Bloom filter is a easy data structure which is space-efficient randomized to bolster enrollment questions. Bloom filters allow false positives. The main objective of this work is to provide information about the ways in which Bloom filters have been used and modified in a variety of network problems, with the point of giving an exceptional scientific and reasonable structure for their utilization in future applications.

3. PROPOSED SYSTEM

In this work, with the focus on privacy-preserving multikeyword similarity search in which all files are encrypted and saved to cloud. Both users and Owners of cloud are registered in the cloud and gains authorization. Upon their access, owners are allowed to send the files along with the keywords attached with to it. Then users Search with their keywords.

Our Problem is more complex than earlier because keywords found in the returned result may have relevant or similar which may differ from input keywords. On the otherhand, our problem is more complex because at a time multiple keywords are taken. As stated earlier, this Multi keyword similarity querycan alsofind numerous applications.

An example for our search scheme is similar to Ib search engine like yahoo, Bing, Google, etc. Because if some of the input keywords are misspelled, still the search engines show the most similar and relevant as possible.

4. CONTRIBUTIONS

Our commitments can be condensed as takes after:

4.1 In this work it initially propose the random traversal calculation which makes the cloud server haphazardly cross on record and returns diverse outcomes for the same inquiry, and meanwhile, it keeps up the exactness of queries unaltered for higher security.

4.2 Based on the Random traversal algorithm, itexhibit one both effective and secure accessible encryption plot, which can bolster top similarity look over encrypted data. In this work, the data owner can control the level of question unlinkability without giving up exactness.

4.3 By the test search technique comes about us to demonstrate that our techniques are more proficient than the cutting edge techniques what's more, can better secure information protection. Particularly, by proposed technique has great adaptability execution when managing vast informational indexes.

5. SYSTEM MODEL

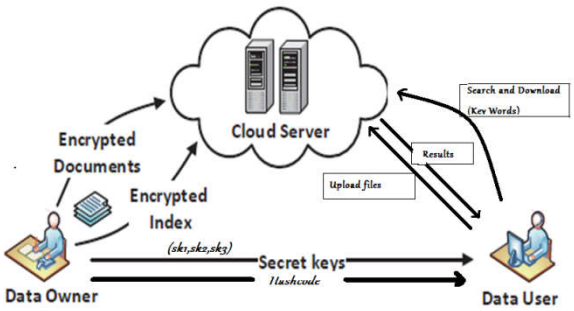

Figure 1. System Architecture

5.1 Design goals:

To empower positioned look for powerful usage of outsourced cloud information under the previously mentioned show, by framework configuration ought to at the similar time accomplish security and execution ensures as takes after.

5.1.1 Multi keyword search: To plan an accessible encryption plot that empowers the cloud server to help multi-keyword similarities investigate encoded information.

5.1.2 Search efficiency: Our plan ought to be proficient in file development, trapdoor age and inquiry handling, furthermore, it ought to be more beneficial and viable than the cutting edge techniques.

5.2 Privacy preserving:

5.2.1. Index security and Query security

5.2.2. Keywords Privacy

5.2.3. Query Unlinkability and Access Pattern

5.3. Background Knowledge

5.3.1 Edit Distance

The edit distance ed(w1, w2) between two words is characterized as the base number of operations required to change starting with single word then onto the next.

5.3.2 Bloom Filter

A Bloom filter is a semi-transient data structure that provides a period and space productive route for testing if a component is a piece of a set. Bloom filters are cool. Consider utilizing one in situations where you require 100% review however a little likelihood of false positives is satisfactory.

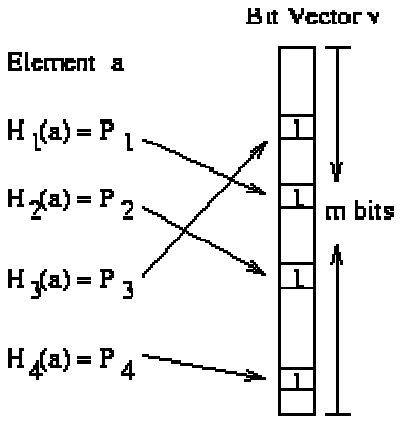

Figure 2. Illustration of Bloom Filter

5.3.2.1 Bloom filter algorithm:

Procedure 1 (IsMember)
IsMember(Table,Key)? Boolean
1. i ? 0
2. repeat
3. i ?i + 1 . hi is the ith hash transform, where 1 <i = m
4. until ((i = m) ? ¬(IsSet(Table[hi(Key)])))
5. if i = m then
6. return(IsSet(Table[hi(Key)]))
7. else
8. return(False)
end.

5.3.3 Random Traversal Algorithm

Random traversal acts correspondingly to randomized breadth first traversal, since regularly the labyrinth must be stretched out without self-crossing on a generally roundabout perimeter. Contrast with depth first traversal. Beginning in the base left corner, the calculation keeps a variety of the conceivable bearings the labyrinth could be broadened. At each progression, the labyrinth is stretched out in one of these arbitrary headings, insofar as doing as such does not reconnect with another piece of the labyrinth.

Require: The query Q, the searchable index I; Ensure: Return k documents with highest scores to the data user;
1: function SEARCH(Q, I, k)
2: for query QCi in query group QC do
3: FINDTOPK(QCi , root of ICi , 0, k)
4: Merge top-k documents listi of QCi into CList
5: end for
6: for document Di in CList do
7: if Score(QR, IRi) > k-th score in Re c then
8: Insert i into Re c
9: end if
10: end for
11: return top-k documents of Re c
12: end function

13:
14: function FINDTOPK(QCi , node, sco, k)
15: if sco< k-th score in listi then
16: return
17: end if
18: if node is leaf node then
19: Insert the f id of node into listi .
20: else
21: leftScore = Score(QCi , node.lc)
22: rightScore = Score(QCi , node.rc)
23: if leftScore>rightScore then
24: FINDTOPK(QCi , node.lc, leftScore, k)
25: FINDTOPK(QCi , node.rc, rightScore, k)
26: else
27: FINDTOPK(QCi , node.rc, rightScore, k)
28: FINDTOPK(QCi , node.lc, leftScore, k) 29: end if
30: end if
31: end function

6. SECURITY ANALYSIS

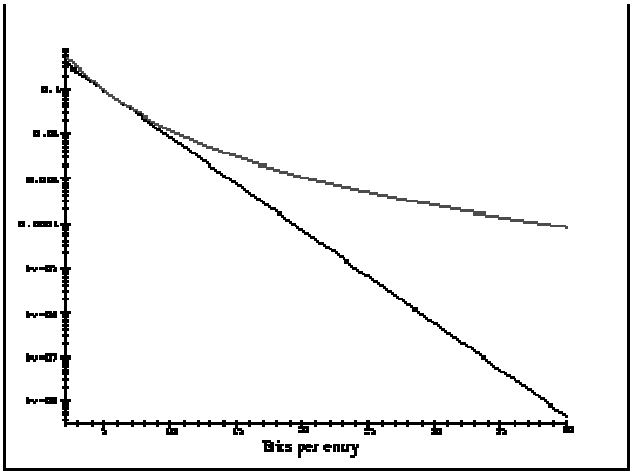

Figure 3. Probability of false positives (log scale).

The best bend is for 4 hash capacities. The base bend is for the ideal (fundamental) number of the hash capacities.

The asymptotic expected greatest tally in the wake of embeddings n keys with k hash capacities into a bit exhibit of size m is

$$\Gamma^{-1}(m)\left(1+\frac{\ln(kn/m)}{\ln\Gamma^{-1}(m)}+O\left(\frac{1}{\ln^2\Gamma^{-1}(m)}\right)\right), \quad (1)$$

Also, the likelihood that any tally is more noteworthy or equivalent I is

$$\Pr(\max(c)\geq i)\leq m\binom{nk}{i}\frac{1}{m^i}\leq m\left(\frac{enk}{im}\right)^i. \quad (2)$$

As of now specified the ideal incentive for k (over reals) is $\ln 2m/n$ assuming to the point that the quantity of hash capacities is short of what this work can additionally bound as

$$\Pr(\max(c) \geq i) \leq m \left(\frac{e \ln 2}{i}\right)^i. \quad (3)$$

Subsequently taking i=16 I acquire that

$$\Pr(\max(c) \geq 16) \leq 1.37 \times 10^{-15} \times m.$$

As such on the off chance that permits 4 bits for every tally, the likelihood of flood for handy estimations of m amid the underlying inclusion in the table is infinitesimal.

7. CONCLUSION

In this work, key-word search center on enhancing the efficiency and the security of multi-keyword top similar seek over encrypted data.

At to begin with, itpropose the Random traversal algorithm which can accomplish that for two indistinguishable queries with various keys, the cloud server crosses unique ways on the list, and the information client gets extraordinary comes about yet with a similar abnormal state of question accuracies meanwhile. At that point, to enhance the pursuit of effectiveness, it design the group multi-keyword top search scheme, which partitions the dictionary into numerous gatherings, and just needs to store the best ck archives of each word assemble when building list.

Next, to secure the query unlink ability, we apply the Random traversal Algorithm to get the RGMTS, which can build the trouble of cloud servers to direct linkage attacks on two indistinguishable queries; Furthermore, this work can likewise tune the estimation of E to make the level of question unlink ability adaptable for data owners. At last, the test comes about demonstrate that search techniques are more effective in efficiency and storage computation and communication overhead.

ACKNOWLEDGEMENTS

The authors would like to thank A. Broder and M. Mitzenmacher, Ning Cao, Cong Wang, Ming Li, Kui Ren, and InjingLou, for throwing light on Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data.

REFERENCES

[1] *Ning Cao, Cong Wang, Ming Li, Kui Ren, and Injing Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data" in IEEE Transactions On Parallel And Distributed Systems, Vol. 25, No. 1, January 2014.*

[2] *J. Sun, R. Zhang, and Y. Zhang, "Privacy-preserving spatialtemporal matching," in Proc. IEEE INFOCOM, 2013, pp. 800–808. [29] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. ICDCS, 2010, pp. 253–262.*

[3] *C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467–1479, Aug. 2012.*

[4] *M. Armbrust et al., "A view of cloud computing," Commun. ACM, vol. 53, no. 4, pp. 50–58, Apr. 2010.*

[5] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advances in Cryptology Eurocrypt 2004. Springer, 2004, pp. 506–522.
[6] K. Ren, C. Wang, Q. Wang et al., "Security challenges for the public cloud," IEEE Internet Computing, vol. 16, no. 1, pp. 69–73, 2012.
[7] L. L. Gremilion. "Designing a Bloom Filter for Differential File Access." Communications of the ACM 25 (1982), 600—604
[8] M. J. Atallah, "Algorithms and Theory of Computation Handbook". Boca Raton, FL, USA: CRC Press, 1998.
[9] A. Broder and M. Mitzenmacher, "Network applications of bloom filters: A survey," Internet Math., vol. 1, pp. 485–509, 2005.
[10] S. Kamara and K. Lauter, "Cryptographic cloud storage," in Proc. Financ. Cryptography Data Secur., 2010, pp. 136–149.
[11] J. Baek, R. Safavi-Naini, and W. Susilo, "Public key encryption with keyword search revisited," in Proc. Int. Conf. Comput. Sci. Appl. (ICCSA), 2008, pp. 1249–1259.