ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

A Multi-keyword Ranked Search Scheme that is Dynamic and Secure over Cloud Data

Priya S [1], Ambika P R [2]

M.Tech Student(IV Sem), Department of Computer Science and Engineering, City Engineering College, Doddakallasandra, Karnataka, India[1]

Assistant Professor, Department of Computer Science and Engineering, City Engineering College, Doddakallasandra, Karnataka, India[2]

ABSTRACT*:* Because of the expanding ubiquity of distributed computing, increasingly information proprietors are propelled to outsource their information to cloud servers for awesome comfort and decreased expense in information administration. Nonetheless, delicate information ought to be encoded some time recently outsourcing for protection prerequisites, which obsoletes information usage like catchphrase based report recovery. In this paper, we show a protected multi-watchword positioned look plan over scrambled cloud information, which all the while bolsters dynamic overhaul operations like cancellation and insertion of reports. In particular, the vector space model and the generally utilized TFIDF model are consolidated as a part of the record development and question era. We build an uncommon tree-based list structure and propose a "Ravenous Profundity first Seek" calculation to give productive multi-catchphrase positioned look. The protected kNN calculation is used to encode the record and question vectors, and in the interim guarantee exact importance score count between scrambled file and question vectors. Keeping in mind the end goal to stand up to measurable assaults, ghost terms are added to the record vector for blinding indexed lists . Because of the utilization of our uncommon tree-based list structure, the proposed plan can accomplish sub-straight pursuit time and manage the erasure and insertion of reports adaptably. Broad tests are led to exhibit the effectiveness of the proposed plan.

KEYWORDS*:* Searchable encryption, multi-keyword ranked search, dynamic update, cloud computing.

I. INTRODUCTION

Distributed computing has been considered as another model of big business IT foundation, which can compose immense asset of processing, stockpiling and applications, empower clients to appreciate universal, advantageous what's more, on-interest system access to a mutual pool of configurable processing assets with awesome productivity, negligible financial overhead. Pulled in by these engaging components, both people and endeavors are propelled to outsource their information to the cloud, rather than acquiring programming and equipment to deal with the information themselves. Notwithstanding of the different points of interest of cloud administrations, outsourcing delicate data, (for example, messages, individual wellbeing records, organization fund information, government reports, and so on.) to remote servers brings protection concerns. The cloud administration suppliers (CSPs) that keep the information for clients might get to clients' delicate data without approval. A general way to deal with secure the information privacy is to scramble the information some time recently outsourcing. Nonetheless, this will bring about a gigantic expense in terms of information ease of use. For instance, the current strategies on catchphrase based data recovery, which are generally utilized on the plaintext information, can't be specifically connected on the encoded information. Downloading all the information from the cloud and decode locally is clearly unfeasible. Keeping in mind the end goal to address the above issue, scientists have composed some universally useful arrangements with completely homomorphic encryption or negligent RAMs. Nonetheless, these strategies are not pragmatic because of their high computational overhead for both the cloud separate and client. Despite what might be expected, more useful specialpurpose arrangements, for example, searchable encryption (SE) plans have made particular commitments regarding effectiveness, usefulness and security. Searchable encryption plans empower the customer to store the

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

scrambled information to the cloud and execute catchphrase look over ciphertext space. As such, plentiful works have been proposed under various danger models to accomplish different hunt usefulness, for example, single watchword inquiry, comparability seek, multi-watchword boolean look, positioned seek, multi-catchphrase positioned look, and so on. Among them, multikeyword positioned seek accomplishes increasingly consideration for its useful materialness. As of late, some dynamic plans have been proposed to bolster embeddings and erasing operations on archive accumulation. These are huge fills in as it is exceedingly conceivable that the information proprietors need to overhaul their information on the cloud server. Be that as it may, few of the dynamic plans support proficient multikeyword positioned look.

This paper proposes a safe tree-based search plan over the scrambled cloud information, which underpins multikeyword positioned inquiry and element operation on the archive gathering. In particular, the vector space model what's more, the broadly utilized "term frequency (TF) × inverse document frequency" model are consolidated in the list development and inquiry era to give multikeyword positioned seek. So as to acquire high inquiry productivity, we build a tree-based file structure and propose an "greedy depth first search" calculation based on this record tree. Because of the unique structure of our tree-based record, the proposed seek plan can adaptably accomplish sub-straight pursuit time and manage the cancellation and insertion of archives. Our commitments are condensed as takes after:

1) We design a searchable encryption scheme that supports both the accurate multi-keyword ranked search and flexible dynamic operation on document collection.
2) Due to the special structure of our tree-based index, the search complexity of the proposed scheme is fundamentally kept to logarithmic. And in practice, the proposed scheme can achieve higher search efficiency by executing our "Greedy Depth-first Search" algorithm. Moreover, parallel search can be flexibly performed to further reduce the time cost of search process.

II. RELATED WORK

Searchable encryption plans empower the customers to store the scrambled information to the cloud and execute watchword seek over ciphertext area. Because of various cryptography primitives, searchable encryption plans can be built utilizing open key based cryptography on the other hand symmetric key based cryptography.

Cloud computing transforms the way information technology (IT) is consumed and managed, promising improved cost efficiencies, accelerated innovation, faster time-to-market, and the ability to scale applications on demand (Leighton, 2009). According to Gartner, while the hype grew exponentially during 2008 and continued since, it is clear that there is a major shift towards the cloud computing model and that the benefits may be substantial (Gartner Hype-Cycle, 2012). However, as the shape of the cloud computing is emerging and developing rapidly both conceptually and in reality, the legal/contractual, economic, service quality, interoperability, security and privacy issues still pose significant challenges. In this chapter, we describe various service and deployment models of cloud computing and identify major challenges. In particular, we discuss three critical challenges: regulatory, security and privacy issues in cloud computing. Some solutions to mitigate these challenges are also proposed along with a brief presentation on the future trends in cloud computing deployment.

Today, cloud computing is being defined and talked about across the ICT industry under different contexts and with different definitions attached to it. The core point is that cloud computing means having a server firm that can host the services for users connected to it by the network. Technology has moved in this direction because of the advancement in computing, communication and networking technologies. Fast and reliable connectivity is a must for the existence of cloud computing. Cloud computing is clearly one of the most enticing technology areas of the current times due, at least in part to its cost-efficiency and flexibility. However, despite the surge in activity and interest, there are significant, persistent concerns about cloud computing that are impeding the momentum and will eventually compromise the vision of cloud computing as a new IT procurement model. Despite the trumpeted business and technical advantages of cloud computing, many potential cloud users have yet to join the cloud, and those major corporations that are cloud users are for the most part putting only their less sensitive data in the cloud.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

III. THE SYSTEM AND THE THREAT MODELS

The system model in this paper involves three different entities: data owner, data user and cloud server, as illustrated in Fig. 1.

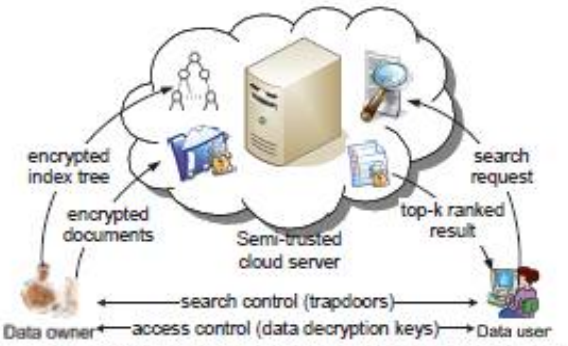

Fig. 1. The architecture of ranked search over encrypted cloud data

Information proprietor has a gathering of archives F = {f1; f2; :::; fn} that he needs to outsource to the cloud server in scrambled structure while as yet keeping the ability to look on them for successful use. In our plan, the information proprietor firstly forms a protected searchable tree record I from archive accumulation F, and afterward creates a scrambled archive gathering C for F. A while later, the information proprietor outsources the scrambled gathering C and the protected record I to the cloud server, what's more, safely appropriates the key data of trapdoor era (counting catchphrase IDF values) and record decoding to the approved information clients. Plus, the information proprietor is in charge of the upgrade operation of his reports put away in the cloud server. While upgrading, the information proprietor produces the overhaul data locally and sends it to the server. Information clients are approved ones to get to the reports of information proprietor. With t question watchwords, the approved client can produce a trapdoor TD concurring to hunt control systems to get k scrambled archives from cloud server. At that point, the information client can decode the reports with the mutual mystery key. Cloud server stores the encoded report accumulation C and the encoded searchable tree file I for information proprietor. After getting the trapdoor TD from the information client, the cloud server executes look over the list tree I, lastly returns the comparing gathering of top- k positioned encoded records. In addition, after getting the overhaul data from the information proprietor, the server necessities to redesign the list I and report gathering C as indicated by the got data. The cloud server in the proposed plan is considered as "fair however inquisitive", which is utilized by loads of takes a shot at secure cloud information look. In particular, the cloud server sincerely and effectively executes directions in the assigned convention. In the mean time, it is inquisitive to construe and dissect got information, which makes a difference it gain extra data. Contingent upon what data the cloud server knows, we embrace the two danger models proposed by Cao et al.

Known Ciphertext Model. In this model, the cloud server just knows the encoded record gathering C, the searchable record tree I, and the inquiry trapdoor TD put together by the approved client. That is to say, the cloud server can direct ciphertext-just assault (COA) in this model.

Known Foundation Model. Contrasted and known ciphertext model, the cloud server in this more grounded model is outfitted with more information, for example, the term recurrence (TF) insights of the record gathering. This factual data records what number of archives are there for every term recurrence of a particular watchword in the entire report gathering, as appeared in Fig. 2, which could be utilized as the watchword character. Outfitted with such factual data, the cloud server can direct TF factual assault to reason or even distinguish certain watchwords through examining histogram and quality extent of the relating recurrence dispersions.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

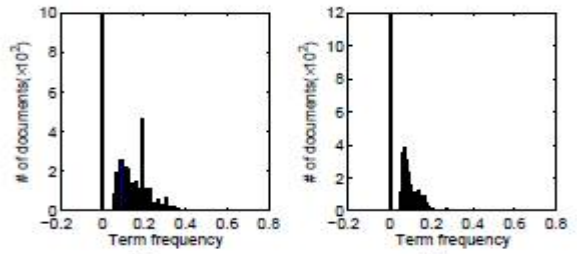

Fig. 2. Distribution of term frequency (TF) for (a) keyword "subnet", and (b) keyword "host".

IV. THE PROPOSED SCHEME

The proposed scheme in this section describes the unencrypted dynamic multi-keyword ranked search (UDMRS) scheme which is constructed on the basis of vector space model and KBB tree. Based on the UDMRS scheme, two secure search schemes (BDMRS and EDMRS schemes) are constructed against two threat models, respectively.

4.1 Index Construction of UDMRS Scheme

We have quickly presented the KBB file tree structure, which helps us in presenting the file development. During the time spent file development, we to start with produce a tree hub for every record in the accumulation. These hubs are the leaf hubs of the file tree. At that point, the inside tree hubs are produced based on these leaf hubs. A case of our record tree is appeared in Fig. 3. Note that the list tree T worked here is a plaintext. Taking after are a few documentations for Calculation 1. Additionally, the information structure of the tree hub is characterized as ⟨ID;D; Pl; Pr; FID⟩, where the interesting personality ID for each tree hub is produced through the capacity GenID().

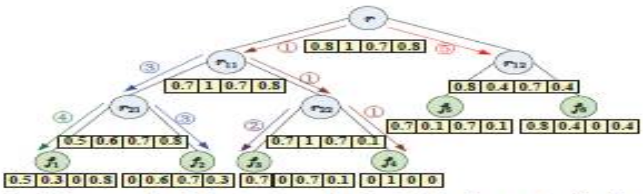

Fig. 3. An example of the tree-based index with the document collection $\mathcal{F} = \{f_i | i = 1, ..., 6\}$ and cardinality of the dictionary $m = 4$. In the construction process of the tree index, we first generate leaf nodes from the documents. Then, the internal tree nodes are generated based on the leaf nodes. This figure also shows an example of search process, in which the query vector Q is equal to $(0, 0.92, 0, 0.38)$. In this example, we set the parameter $k = 3$ with the meaning that three documents will be returned to the user. According to the search algorithm, the search starts with the root node, and reaches the first leaf node f_4 through r_{11} and r_{22}. The relevance score of f_4 to the query is 0.92. After that, the leaf nodes f_3 and f_2 are successively reached with the relevance scores 0.038 and 0.67. Next, the leaf node f_1 is reached with score 0.58 and replace f_3 in $RList$. Finally, the algorithm will try to search subtree rooted by r_{12}, and find that there are no reasonable results in this subtree because the relevance score of r_{12} is 0.52, which is smaller than the smallest relevance score in $RList$.

• CurrentNodeSet – The arrangement of current preparing hubs which have no folks. On the off chance that the quantity of hubs is even, the cardinality of the set is signified as 2h(h ∈ Z+), else the cardinality is signified as (2h + 1).

• TempNodeSet – The arrangement of the recently produced hubs. In the file, if Du[i]≠ 0 for an inner hub u, there is no less than one way from the hub u to some leaf, which shows a report containing the catchphrase wi. Moreover Du[i] dependably stores the greatest standardized TF estimation of wi among its kid hubs. Along these lines, the conceivable Biggest relevance score of its children can be easily estimated.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

4.2 Search Process of UDMRS Scheme

The inquiry procedure of the UDMRS plan is a recursive technique upon the tree, named as "Greedy Depthfirst Search (GDFS)" calculation. We build an outcome list indicated as RList, whose component is characterized as ⟨RScore; FID⟩. Here, the RScore is the importance score of the record fFID to the inquiry, which is figured as indicated by Equation (1). The RList stores the k got to reports with the biggest significance scores to the question.

Algorithm GDFS(IndexTreeNode *u*)
1: **if** the node *u* is not a leaf node **then**
2: **if** RScore(*Du;Q*) > *kthscore* **then**
3: GDFS(*u:hchild*);
4: GDFS(*u:lchild*);
5: **else**
6: **return**
7: **end if**
8: **else**
9: **if** RScore(*Du;Q*) > *kthscore* **then**
10: Delete the element with the smallest relevance score from *RList*;
11: Insert a new element *⟨*RScore(*Du;Q*)*; u:FID⟩* and sort all the elements of *RList*;
12: **end if**
13: **return**
14: **end if**

V. PERFORMANCE ANALYSIS

The search precision of scheme is affected by the dummy keywords in EDMRS scheme. Here, the 'precision' is defined as that in [26]: $Pk = k'=k$, where k' is the number of real top-k documents in the retrieved k documents. If a smaller standard deviation _ is set for the random variable Σ *"v*, the EDMRS scheme is supposed to obtain higher precision, and vice versa. The results are shown in Fig. 4(a).

In the EDMRS scheme, phantom terms are added to the index vector to obscure the relevance score calculation, so that the cloud server cannot identify keywords by analyzing the *TF* distributions of special keywords. Here, we quantify the obscureness of the relevance score by "rank privacy", where *ri* is the rank number of document in the retrieved top-*k* documents, and *r′ i* is its real rank number in the whole ranked results. The larger rank privacy denotes the higher security of the scheme, which is illustrated in Fig. 4(b).

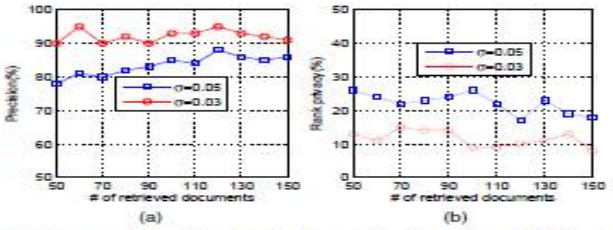

Fig. 4. The precision (a) and rank privacy (b) of searches with different standard deviation σ.

VI. CONCLUSION

In this paper, a safe, productive and dynamic pursuit plan is proposed, which bolsters not just the exact multi-catchphrase positioned seek additionally the element cancellation and insertion of records. We build an exceptional catchphrase adjusted double tree as the list, what's more, propose a "Covetous Profundity first Inquiry" calculation to

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

get preferable effectiveness over straight inquiry. Furthermore, the parallel pursuit procedure can be completed to promote diminish the time cost. The security of the plan is ensured against two danger models by utilizing the protected kNN calculation. Trial results show the proficiency of our proposed plan. There are still numerous test issues in symmetric SE plans. In the proposed plan, the information proprietor is in charge of creating overhauling data and sending them to the cloud server. In this way, the information proprietor requirements to store the decoded record tree and the data that are important to recalculate the IDF values. It is not practical and a dishonest data user will lead to many secure problems. For example, a dishonest data user may search the documents and distribute the decrypted documents to the unauthorized ones. Even more, a dishonest data user may distribute his/her secure keys to the unauthorized ones. In the future works, we will try to improve the SE scheme to handle these challenge problems.

REFERENCES

[1] 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.
[2] S. Kamara and K. Lauter, "Cryptographic cloud storage," in *Financial Cryptography and Data Security*. Springer, 2010, pp. 136–149.
[3] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford University, 2009.
[4] O. Goldreich and R. Ostrovsky, "Software protection and simulation on oblivious rams," *Journal of the ACM (JACM)*, vol. 43, no. 3, pp.
[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] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. E. Skeith III, "Public key encryption that allows pir queries," in *Advances in Cryptology-CRYPTO 2007*. Springer, 2007, pp. 50–67.
[7] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on*. IEEE, 2000, pp. 44– 55.
[8] E.-J. Goh *et al.*, "Secure indexes." *IACR Cryptology ePrint Archive*, vol. 2003, p. 216, 2003.
[9] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in *Proceedings of the Third international conference on Applied Cryptography and Network Security*. Springer-Verlag, 2005, pp. 442–455.
[10] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in *Proceedings of the 13th ACM conference on Computer and communications security*. ACM, 2006, pp. 79–88.
[11] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010, pp. 1–5.
[12] M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," in *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*. IEEE, 2012, pp. 1156–1167.
[13] C. Wang, K. Ren, S. Yu, and K. M. R. Urs, "Achieving usable and privacy-assured similarity search over outsourced cloud data," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 451–459. 1045-9219 (c) 2015 IEEE.
[14] B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud," in *IEEE INFOCOM*, 2014.