© 2019 IJSRST | Volume 6 | Issue 2 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology
DOI : 10.32628/IJSRST19624

Secure and Dynamic Multi-Keyword Ranked Search Scheme Over Encrypted Cloud Data Using K-Means Clustering

Tarika P. Jawale, Prof. R. B. Mapari
G. S. Mandal's MIT, Aurangabad, Maharashtra, India

ABSTRACT

A Secure and Dynamic Multi-keyword graded Search theme over Encrypted Cloud information attributable to the increasing fame of cloud computing, a lot of information homeowners are spurred to source their information to cloud servers for unimaginable accommodation and diminished expense in information management can also perform information dynamic operations on files. On the opposite hand, sensitive information needs to be encrypted before outsourcing for security conditions, that obsoletes information use like keyword-based document retrieval. A protected multi-keyword graded search theme over encrypted cloud information, that all the whereas underpins part update operations like deletion and insertion of documents. Especially, the vector area model and therefore the usually utilised TF_IDF model are consolidated as a neighbourhood of the index development and question generation. A unique tree-based index structure employing a "K-means Clustering" formula to provide practiced multi-keyword graded search. The secure KNN formula is employed to cipher the index and question vectors, so guarantee precise importance score calculation between encrypted index and question vectors. With a selected finish goal to oppose measurable attacks, phantom terms are accessorial to the index vector for glaring search results. Due to the employment of our exceptional tree-based index structure. Keyword: Reduplication, Authorized duplicate check, public auditing, shared data, Cloud computing.

Keywords : K-Means Clustering, KNN, Cloud Computing, Inverse Document Return, EDMRS, Searchable Encryption, Multi-Keyword Ranked Search, Dynamic Update

I. INTRODUCTION

CLOUD computing has been thought-about as another model of enterprise IT infrastructure, which may compose mammoth resource of computing, storage and applications, and empower users to understand pervasive, useful and on demand network access to a mutual pool of configurable computing resources with unbelievable potency and insignificant economic overhead. Force in by these partaking options, each people and enterprises are roused to source their information to the cloud, instead of shopping for software system and hardware to influence the information themselves. In spite of the various points of interest of cloud services, outsourcing delicate info, (for example, e-mail, individual health records, organization account info, government archives, so forth.) to remote servers brings privacy considerations. The cloud service suppliers (CSPs) that keep for users might access users' sensitive information while not authorization.

Inverse document return (IDF)" model are joined within the list development and inquiry era to allow multi keyword positioned ask for. Keeping in mind the tip goal to induce high search Effectiveness, we

have a tendency to develop a tree primarily based list structure victimisation a "K-means Clustering" calculation supported this list tree. Thanks to the uncommon structure of our tree-based list, the search theme will flexibly accomplish sub-straight search time and manage the deletion and insertion of reports. The protected KNN formula is employed to encipher the index and question vectors, and within the interim guarantee connection score calculation between encrypted index and question vectors. To oppose distinctive attacks in several threat models, we have a tendency to build 2 secure search themes: the fundamental dynamic multi-keyword hierarchic search (BDMRS) scheme within the acknowledged cipher text model, and therefore the increased dynamic multi-keyword hierarchic search (EDMRS) theme within the acknowledged background model. Our commitments are condensed as takes after:

1) Style a searchable cryptography theme that underpins each the precise multi-keyword hierarchic search and versatile dynamic operation on document assortment for multiple information owner atmospheres.
2) Because of the uncommon structure of our tree-based index, the search quality is during a general sense unbroken to index. What is additional, much speaking; the theme will accomplish higher search proficiency by capital punishment our "K-means clustering" formula. In addition, parallel search may be flexibly performed to more reduce the time value of inquiry procedure.

II. LITERATURE SURVEY

1] Fuzzy Keyword Search over Encrypted Data in Cloud Computing

Authors: Jaydip Sen

In this paper, for the first time an inclination to formalize and solve the matter of effective fuzzy keyword search over encrypted cloud data whereas maintaining keyword privacy. Fuzzy keyword search greatly enhances system usability by returning the matching files once users' looking inputs specifically match the predefined keywords or the best achievable matching files supported keyword similarity linguistics, once precise match fails. In our answer, we've got an inclination to take advantage of edit distance to quantify keywords similarity and develop an advanced technique on constructing fuzzy keyword sets, that greatly reduces the storage and illustration overheads. Through rigorous security analysis, we've got an inclination to point out that our projected answer is secure and privacy-preserving, whereas properly realizing the goal of fuzzy keyword search easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

2] Practical Techniques for Searches on Encrypted Data

Authors: Seny Kamara

In this paper, an inclination to explain our science schemes for the matter of betting on encrypted data and provide proofs of security for the following crypto systems. Our techniques have form of crucial blessings. they are demonstrably secure: they provide obvious secrecy for cryptography, among the sense that the un trusted server cannot learn one thing relating to the plaintext once only given the cipher text; they provide question isolation for searches, meaning that the un trusted server cannot learn one thing plenty of relating to the plaintext than the search result; they provide controlled trying, that the un trusted server cannot hunt for AN arbitrary word whereas not the user's authorization; they to boot support hidden queries, that the user might raise the un trusted server to travel longing for a secret word whereas not revealing the word to the server.

3] A fully homomorphic encryption scheme

Authors: Reza Carmela, Juan Garay

"In this Paper propose the first fully homomorphic coding theme, taking care of a focal open issue in cryptography. Such a thought permits one to work subjective capacities over encrypted information while not the secret writing key – i.e., given encryptions E (m1),..., E (m t) of M1,...,m t, one will anciently method a smaller cipher text that encrypts f(m1,...,m t) for any efficiently computable capability f. This issue was postured by Rivets et al. in 1978. Fully homomorphism coding has varied applications. as an example, it empowers non-public queries to a hunt engine– the user presents Associate in Nursing encrypted question and therefore the computer programme processes a short-encrypted answer whereas ne'er taking a goose at the question within the clear. It likewise empowers betting on encrypted information – a user stores encrypted field on an overseas file server and may later have the server recover simply files that (when decoded) fulfil a number of Boolean limitations, despite the very fact that the server cannot unscramble the files on their lonesome. All the additional comprehensively, fully homomorphism coding enhances the efficiency of secure multi keyword rank search.

4] Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions

Authors: Dan Bones

In this paper we've an inclination to point out two solutions to south southeast that at a similar time relish the following properties:

1. Every solution are lots of economical than all previous constant round schemes. Specially, the work performed by the server per came back document is constant as hostile linear at intervals the scale of the information.

2. Every solution relish stronger security guarantee than previous constant-round schemes. In fact, we've an inclination for instance delicate but serious problems with previous notions of security for south southeast, and show the thanks to vogue constructions that avoid these pitfalls. Further, our second answer in addition achieves what we've an inclination to call adaptative south southeast security, where queries to the server could also be chosen adaptively (by the adversary) throughout the execution of the search.

III.EXISTING SCHEME

UDMRS Scheme

More formally, an attribute-based cloud data integrity auditing protocol consists of the following six algorithms.

The search process of the UDMRS scheme is a recursive procedure upon the tree, named as “Greedy Depth-first Search” algorithm. We construct a result list denoted as RList, whose element is defined as hRScore; FIDi. Here, the RScore is the relevance score of the document fFID to the query, which is calculated according to Formula (1). The RList stores the k accessed documents with the largest relevance scores to the query. The elements of the list are ranked in descending order according to the RScore, and will be updated timely during the search process. Following are some other notations, and the GDFS algorithm is described in Algorithm 2.

Algorithm 2. GDFS (IndexTreeNode u)

Step 1 : If node a is not a leaf node then

Step 2 : GDFS(lchild);

Step 3 : else

Step 4 : return

Step 5 : end if

Step 6 : else

Step 7 : if Rscore > kth score then

Step 8 : Delete the element with the smallest relevance score from RList;
Step 9 : Insert a new element Rscore and sort all the elements of RList;
Step 10 : return
Step 11: end if

BDMRS Scheme

Based on the UDMRS scheme, we construct the basic dynamic multi-keyword ranked search scheme by using the secure kNN algorithm [8]. The BDMRS scheme is designed to achieve the goal of privacy-preserving in the known ciphertext model, and the four algorithms included are described as follows:
Security analysis. We analyze the BDMRS scheme accord-ing to the three predefined privacy requirements in the design goals:
Index confidentiality and query confidentiality. In the proposed BDMRS scheme, Iu and TD are obfuscated vectors, which means the cloud server cannot infer the original vectors Du and Q without the secret key set SK. The secret keys M1 and M2 are Gaussian random matrices. According to [38], the attacker (cloud server) of COA cannot calculate the matrices merely with cipher text. Thus, the BDMRS scheme is resilient against ciphertext only attack and the index confidentiality and the query confidentiality are well protected.
Query unlink ability. The trapdoor of query vector is generated from a random splitting operation, which means that the same search requests will be trans-formed into different query trapdoors, and thus the query unlinkability is protected. However, the cloud server is able to link the same search requests according to the same visited path and the same relevance scores.

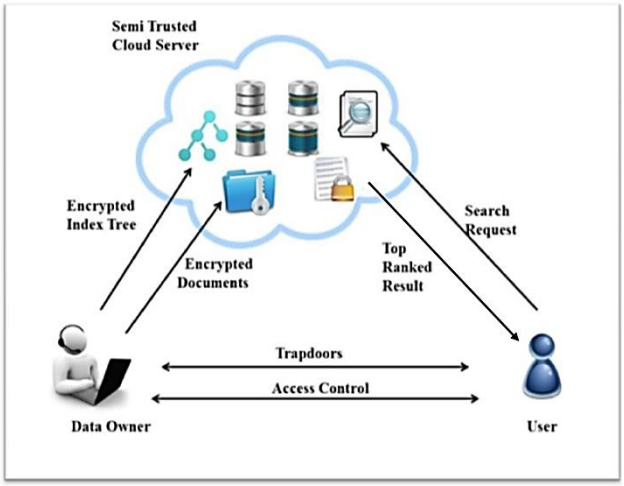
Drawbacks of existing scheme

- This system is more time consuming because of they search one by one node in the index tree.
- Time complexity: not suitable for large data sets.
- There is no guarantee of finding the Target node.

IV. PROPOSED SCHEME

The system model during this paper incorporates 3 clear substances: information homeowners, information user and cloud server, as illustrated in Fig. 1 There square measure multiple information owner in system As information owner incorporates a gathering of records F = That he has to source to the cloud server in encoded structure whereas up 'til currently keeping the power to envision on them for convincing utilization. information owner first manufactures a secure searchable tree index I from archive accumulation F, and a brief time later makes a encrypted document gathering C for F. a quick span later, the information owner outsources the encoded accumulation C and therefore the secure index I to the cloud server, and safely disseminates the key information of trapdoor era and document decoding to the approved information users. To boot, the information owner is aware of his documents hold on within the cloud server. Whereas change, owner creates the upgrade information regionally and sends it to the server can also perform data dynamic operations on files. Data users square measure approved ones to induce to the archives of information owner. With t question keywords, the approved user will produce a trapdoor TD as indicated by search management mechanisms to induce k encrypted documents from cloud server. By then, decipher the documents with the shared secret key. Cloud server stores the encrypted document accumulation C and therefore the encrypted searchable tree index I for information owner.

Within the wake of tolerating the trapdoor TD from the information user, look over the index tree I, lastly provides back the relating gathering of top-k settled encoded reports. Also, within the wake of tolerating the update data from the information owner, the server has to update the index I and document gathering C as per the received data. After insertion or deletion of a record, we have a tendency to need change synchronously the index. Since the index of DMRS theme is planned as a balanced binary tree, the dynamic operation is finished by redesigning hubs within the list tree. The report on record is simply in sight of archive acknowledges, and no entrance to the substance of records is needed.



System architecture

BDMRS Scheme:

In view of the UDMRS scheme, we build the essential element multi-keyword ranked search (BDMRS) scheme by utilizing the secure kNN algorithm. The BDMRS scheme is intended to accomplish the objective of privacy preserving in the known cipher text model. BDMRS scheme can secure the Index Confidentiality and Query Confidentiality in the known cipher text model.

EDMRS Scheme:

Cloud server has the capacity interface the same search requests by following way of visited nodes.The Cloud server recognize a keyword as the standardized TF distribution of the keyword can be precisely acquired from the last computed relevance scores. A heuristic strategy to further enhance the security is to break such correct quality. Hence, we can acquaint some tunable haphazardness with exasperate the significance score estimation. Likewise, to suit diverse users' inclinations for higher exact positioned results or better protected keyword privacy, the arbitrariness are set movable.

Dynamic Update Operation of DMRS:

After insertion or deletion of a record, we require to update synchronously the index. Since the index of DMRS scheme is planned as a balanced binary tree, the dynamic operation is done by redesigning hubs in the list tree. The report on record is just in view of archive recognizes, and no entrance to the substance of records is required.

ALGORITHM FOR SEARCHING:

Step 1 : Clusters the data into k groups where k is predefined.

Step 2 : Select k points at random as cluster canters.

Step 3 : Assign objects to their closest cluster canter according to the Euclidean distance function.

Step 4 : Calculate the centroid or mean of all objects in each cluster.

Step 5 : Repeat steps 2, 3 and 4 until the same points are assigned to each cluster in consecutive rounds.

Advantages of Proposed scheme

- It takes less time for searching process.
- It solves the problem even data set is large.
- It gives accurate searching results.

- It is flexible.

V. RESULTS AND DISCUSSION

We implement the propose scheme using java language in windows7 operation system and test its efficiency on real-world document collection. We compare our scheme with "Greedy Depth First Search" algorithm. Our scheme can also carry out the update operation without storing the index tree on data owner side. The experimental result proved that our scheme will obtain better search efficiency.

Sr. No	Existing scheme	proposed scheme
1	88%	89%
2	94%	92%
3	97%	98.3%
4	100%	95%
5	85%	82%
6	89%	88%
7	89%	86%
8	96%	72.5%
9	96%	86%
10	86.7%	79.7%
11	87.5%	78%
12	100%	92%
13	82%	95.6%
14	100%	100%
15	100%	98%
16	71.1%	93%
Avg. Precision =91%		Avg. Precision =93%

Table Comparison precision values

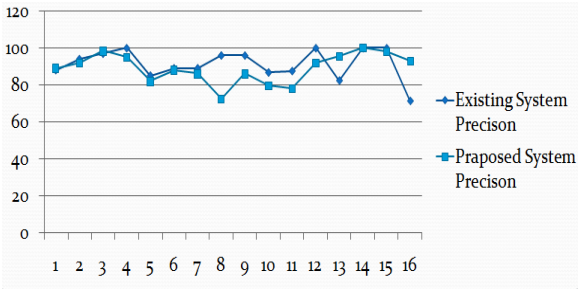


Figure 1. Graph for precision values

Table 1. Performance for searching files

Size/time	Greedy	K Means
10KB	0.9	0.7
20KB	1.6	1.4
40KB	3.2	2.8
80KB	6.2	4.9

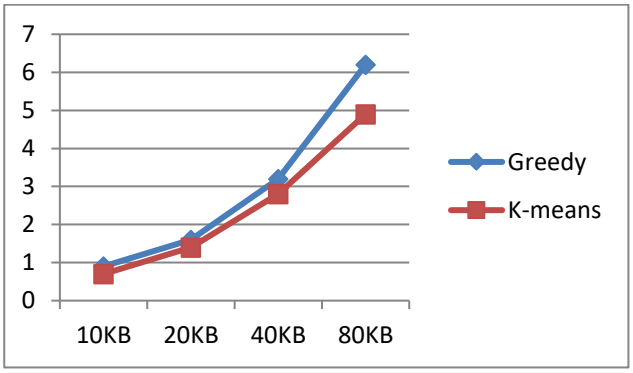


Figure 2. Graph for searching files

VI. CONCLUSION

The Proposed system display's accurate multi keyword ranked search result. Result will be displayed in the ranked format. Useful for perform dynamic operations. Existing system's precision ratio is 91% and proposed system's precision ratio is 93%. Searching time reduced by 3 sec in proposed scheme. For ex. 10kb file taking 0.9 sec in existing scheme where as proposed scheme it takes only 0.6 sec.

VII. REFERENCES

[1] 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.

[2] P. Galle, J. Stardom, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Applied Cryptography and Network Security. Springer, 2004, pp. 31–45.

[3] Y. H. Hwang and P. J. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in Proceedings of the First international conference on Pairing-Based Cryptography.

[4] L. Ballard, S. Karama, and F. Monroe, "Achieving efficient conjunctive keyword searches over encrypted data," in Proceedings of the 7th international conference on Information and Communications Security.

[5] D. Bones and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in Proceedings of the 4th conference on Theory of cryptography. Springer-Vorlage, 2007, pp. 535–554.

[6] B. Zhang and F. Zhang, "An efficient public key encryption with conjunctive-subset keywords search," Journal of Network and Computer Applications, vol. 34, no. 1, pp. 262–267, 2011.

[7] J. Katz, A. Sahai, and B. Waters, "Predicate encryption supporting disjunctions, polynomial equations, and inner products," in Advances in Cryptology–EUROCRYPT 2008. Springer, 2008, pp. 146–162.

[8] E. Shan, E. Shi, and B. Waters, "Predicate privacy in encryption systems," in Proceedings of the 6th Theory of Cryptography Conference on Theory of Cryptography. Springer V erlang, 2009, pp. 457–473.

[9] A. Lawks, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully secure functional encryption: attribute-based encryption and (hierarchical) inner product encryption," in Proceedings of the 29th Annual international conference on Theory and Applications of Cryptographic Techniques. Springer-Vorlage, 2010, pp. 62–91.

[10] Ankit Lodha, Clinical Analytics – Transforming Clinical Development through Big Data, Vol-2, Issue-10, 2016

[11] Ankit Lodha, Agile: Open Innovation to Revolutionize Pharmaceutical Strategy, Vol-2, Issue-12, 2016

[12] Ankit Lodha, Analytics: An Intelligent Approach in Clinical Trail Management, Volume 6, Issue 5 , 1000e124

Cite this article as :

Tarika P. Jawale, Prof. R. B. Mapari, "Secure and Dynamic Multi-Keyword Ranked Search Scheme Over Encrypted Cloud Data Using K-Means Clustering", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 6 Issue 2, pp. 230-235, March-April 2019. Available at doi : https://doi.org/10.32628/IJSRST196238 Journal URL : http://ijsrst.com/IJSRST196238