Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 12, December -2017

A SCRIPT TOWARDS SECURITY-PRESERVING AND NUMEROUS WORD SEARCH METHOD ON ENCRYPTED DATA IN CLOUD COMPUTING PREMISES

[*]NARAHARI AJMEERA, *[*]SANDHYA RANI PABBATHI

[*]*(Department of CSE, JNTUH college of Engineering Manthani, India)*
[**] *(Department of CSE, Joginapally Engineering college, India)*

ABSTRACT: *Cloud information owners prefer to outsource documents in an encrypted form for the motive of private retaining. Therefore it's far critical to expand green and reliable ciphertext search techniques. One project is that the relationship between files may be usually concealed within the system of encryption, so that you can cause enormous search accuracy performance degradation. Also, the extent of statistics in statistics centers has experienced a dramatic increase. This will make it even more hard to design ciphertext seek schemes that may provide green and reliable online information retrieval on a large quantity of encrypted statistics. In this paper, a hierarchical clustering approach is proposed to assist greater seek semantics and also to satisfy the demand for fast ciphertext search inside a big data environment. The proposed hierarchical technique clusters the files primarily based at the minimal relevance threshold and then partition the resulting clusters into sub-clusters until the constraint at the most length of the cluster is reached. In the search section, this method can attain a linear computational complexity in opposition to an exponential size boom of file series. In order to confirm the authenticity of seek results; a structure called minimal hash sub-tree is designed in this paper. Experiments had been performed using the collection set constructed from the IEEE Xplore. The outcomes show that with a sharp boom of files in the dataset the search time of the proposed technique will increase linearly whereas the hunt time of the traditional technique will increase exponentially. Furthermore, the proposed method has a bonus over the traditional method inside the rank private and relevance of retrieved files.*

Keywords: *Cloud computing, ciphertext search, ranked search, multi-keyword search, hierarchical clustering, big data, security*

I. INTRODUCTION

As we step into the huge records era, terabyte of statistics are produced worldwide in step with day. Enterprises and users who own a large amount of records commonly select to outsource their precious facts to cloud facility with the intention to lessen facts control fee and garage facility spending. As an end result, data quantity in cloud storage facilities is experiencing a dramatic growth. Although cloud server providers (CSPs) declare that their cloud carrier is armed with sturdy security features, safety and private are most important limitations preventing the broader attractiveness of cloud computing carrier [1]. A traditional manner to lessen records leakage is facts encryption. However, this will make server-facet records usage, consisting of looking on encrypted information; grow to be a completely difficult task. In the recent years, researchers have proposed many ciphertext search schemes [35-38] [43] via incorporating the cryptography techniques. These techniques were demonstrated with provable safety; however their techniques want big operations and have excessive time complexity. Therefore, former methods aren't appropriate for the huge records situation where data extent may be very large and applications require on line information processing. In addition, the relationship among documents is hid in the above strategies. The dating among files represents the homes of the documents and as a result maintaining the relationship is critical to completely express a document. For instance, the connection can be used to specific its class. If a report is impartial of some other documents except those documents which might be associated with sports activities, then it is simple for us to say this report belongs to the class of the sports activities. Due to the blind encryption, this vital belonging has been hid in the conventional methods. Therefore, providing a way which can maintain and make use of this relationship to speed the hunt segment is suitable. On the opposite hand, due to software program/hardware failure, and storage corruption, facts seek consequences returning to the customers may additionally include broken data or were distorted by means of the malicious administrator or intruder. Thus, a verifiable mechanism should be furnished for users to verify the correctness and completeness of the quest effects.

II. RELATED WORK

In recent years, searchable encryption which provides text search function based on encrypted data has been widely studied, especially in security definition, formalizations and efficiency improvement, e.g. [2-7]. As shown in Fig.1, the proposed method is compared with existing solutions and has the advantage in maintaining the relationship between documents.

Single Keyword Searchable Encryption: Song et al [8] first introduced the belief of searchable encryption. They endorse to encrypt each word within the document independently. This approach has an excessive searching value because of the scanning of the whole data series phrase via word. Goh et al [9] officially described at ease index structure and formulate a safety version for index known as semantic protection against adaptively chosen keyword attack (ind-cka). They additionally developed an efficient ind-cka relaxed index construction called z-idx by using the use of pseudo-random capabilities and bloom filters. Cash et al [2] currently layout and enforce an efficient data shape. Due to the dearth of rank mechanism, customers must take a long term to choose what they want while massive files comprise the query key-word. Thus, the order-preserving strategies are applied to recognize the rank mechanism. Wang et al [3] use encrypted invert index to attain comfy ranked keyword seek over the encrypted files. In the quest segment, the cloud server computes the relevance rating among files and the question. In this way, applicable documents are ranked in line with their relevance rating and customers can get the pinnacle-okay results. In the public key putting, Boneh et al [3] designed the primary searchable encryption production, wherein everybody can use the public key to jot down to the statistics stored on the server but most effective authorized users proudly owning private key can seek. However, all the above-mentioned strategies handiest support unmarried key-word search.

Multiple Keyword Searchable Encryptions: To enhance search predicates, a variety of conjunctive keyword search strategies had been proposed. These strategies display huge overhead, which includes communication price via sharing a mystery, eg. [5] or computational value with the aid of bilinear map, e.g. [7]. Pang et al [8] recommend a cozy search scheme based totally on vector space version. Due to the dearth of the security evaluation for frequency facts and realistic search performance, it's far doubtful whether or not their scheme is cozy and green or not. Cao et al [9] present a unique architecture to remedy the problem of multi-keyword ranked seek over encrypted cloud records. But the hunt time of this method grows exponentially accompanying with the exponentially increasing size of the document collections. Sun et al [6] give a new architecture which achieves better search performance. However, at the stage of index constructing process, the relevance between files is omitted. As an end result, the relevance of plaintexts is hid by means of the encryption, customers expectation can't be fulfilled well. For instance: given a question containing Mobile and Phone, only the documents containing each of the key phrases may be retrieved with the aid of conventional methods. But if taking the semantic dating between the documents into consideration, the files containing Cell and Phone must also be retrieved. Obviously, the second one end result is better at assembly the consumer's expectation.

III. TECHNIQUES IMPLEMENTED

In this paper, we propose a multi-key-word ranked seek over encrypted records based on hierarchical clustering index (MRSE-HCI) to keep the near relationship between unique simple files over the encrypted domain so that you can beautify the quest efficiency. In the proposed architecture, the hunt time has a linear growth accompanying with an exponential growing size of information series. We derive this idea from the observation that users retrieval needs usually deal with a particular discipline. So we are able to accelerate the searching procedure by means of computing relevance score among the question and documents which belong to the equal particular subject with the query As a result, only files which might be categorized to the field specified via customers query can be evaluated to get their relevance rating. Due to the irrelevant fields omitted, the quest velocity is stronger. We look into the hassle of keeping the close relationship between unique simple documents over an encrypted domain and endorse a clustering approach to resolve this trouble. According to the proposed clustering approach, every report may be dynamically classified into a particular cluster which has a constraint at the minimum relevance rating among one of a kind documents inside the dataset. The relevance score is a metric used to assess the relationship among distinctive files. Due to the new documents introduced to a cluster, the constraint at the cluster may be damaged. If one of the new documents breaks the constraint, a brand new cluster center could be added and the cutting-edge record could be selected as a temporal cluster center. Then all of the documents can be reassigned and all of the cluster facilities may be reelected. Therefore, the variety of clusters depends at the range of files inside the dataset and the near dating among one of a kind undeniable files. In different phrases, the cluster facilities are created dynamically and the number of clusters is decided by means of the assets of the dataset. We endorse a hierarchical method with the intention to get a better clustering end result within a massive amount of statistics series. The length of every cluster is controlled as a trade-off among clustering accuracy and question efficiency. According to the proposed technique, the number of clusters and the minimal relevance score growth with the boom of the stages while the most length of a cluster reduces. Depending at

the needs of the grain stage, the most size of a cluster is about at each stage. Every cluster wishes to satisfy the restrictions. If there's a cluster whose size exceeds the problem, this cluster may be divided into several sub-clusters.

System Model: The system model contains three entities, as illustrated in Fig. 1, the data owner, the data user, and the cloud server. The box with dashed lines in the figure indicates the added component to the existing architecture.

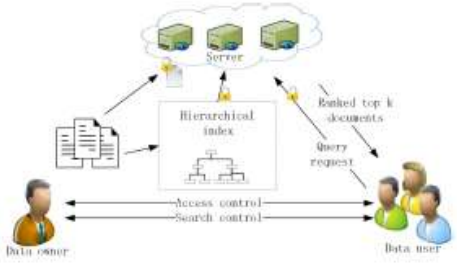
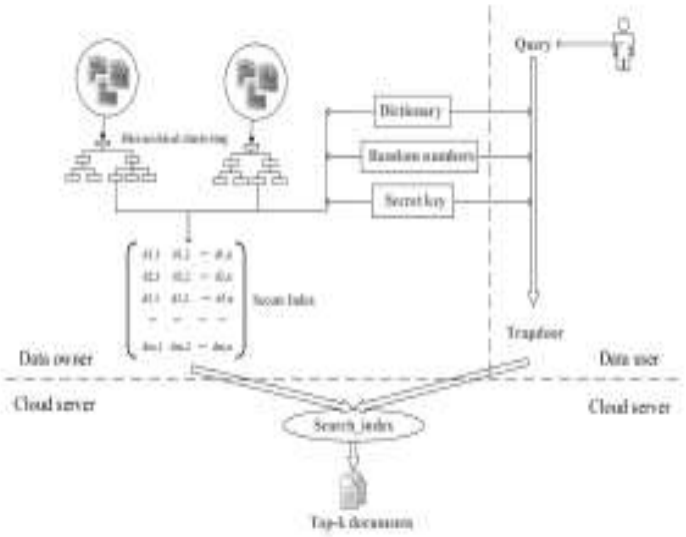


Fig: 1 system architecture

The statistics proprietor is liable for collecting documents, building file index and outsourcing them in an encrypted layout to the cloud server. Apart from that, the information consumer wishes to get the authorization from the information owner before access to the statistics. The cloud server offers a big garage area, and the computation resources needed by ciphertext search. Upon receiving a legal request from the statistics consumer, the cloud server searches the encrypted index and sends back top-k documents which are most probably to healthy users question [12]. The quantity ok is well chosen by using the statistics user. Our device aims at shielding information from leaking data to the cloud server even as improving the performance of ciphertext seek. In this version, each the facts proprietor and the information person are relied on, at the same time as the cloud server is semi-relied on, that is consistent with the architecture in [10, 19, and 29]. In different phrases, the cloud server will strictly observe the predicated order and try to get more records about the information and the index.

MRSE-HCI Architecture: MRSE-HCI architecture is depicted by Fig. 2, where the data owner builds the encrypted index depending on the dictionary, random numbers and secret key, the data user submits a query to the cloud server for getting desired documents, and the cloud server returns the target documents to the data user.



In this segment, we can introduce the MRSE-HCI scheme. The vector space model followed by way of the MRSE-HCI scheme is same because the MRSE [19], while the technique of constructing index is completely distinct. The hierarchical index structure is brought into the MRSE-HCI in place of sequence index. In MRSE-HCI, each record is indexed by a vector. Every size of the vector stands for a keyword and the value represents whether or not the keyword seems or not within the document. Similarly, the question is likewise represented by means of a vector. In the search section, cloud server calculates the relevance rating between the question and documents by computing the inner fabricated from the query vector and file vectors and returns the target documents to the person in step with the pinnacle of relevance score. Due to the fact that each one the documents outsourced

Quality Hierarchical Clustering Algorithm: So a ways, a variety of hierarchical clustering techniques have been proposed. However, all of these techniques are not akin to the partition clustering approach in terms of time complexity overall

performance. K-means [31] and K-medias are famous partition clustering algorithms. But the okay is fixed in the above two methods, which can't be applied to the situation of a dynamic variety of cluster centers. We propose a satisfactory hierarchical clustering (QHC) set of rules primarily based on the unconventional dynamic K-method. As the proposed dynamic K-method algorithm shown inside the Fig. Four, the minimal relevance threshold of the clusters is defined to keep the cluster compact and dense. If the relevance rating between a file and its middle is smaller than the brink, a brand new cluster center is brought and all of the files are reassigned. The above manner could be iterated until ok is solid. Comparing with the traditional clustering approach, ok is dynamically modified during the clustering method. This is why it is referred to as dynamic K-method set of rules.

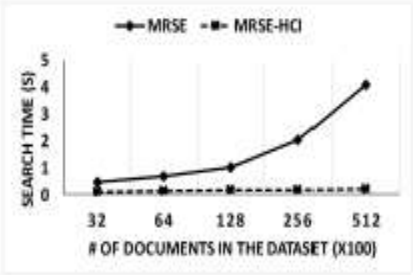


Fig: search time with increasing documents

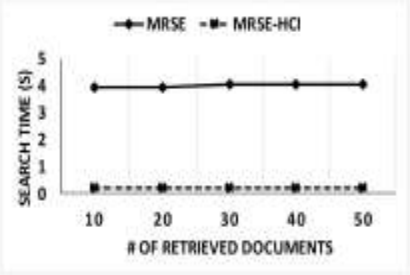


Fig: Search time with the increasing number of retrieved documents

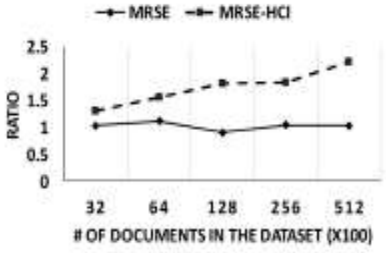


Fig: Relevance of documents

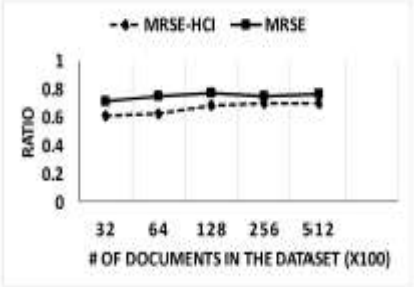


Fig: Relevance between documents and query

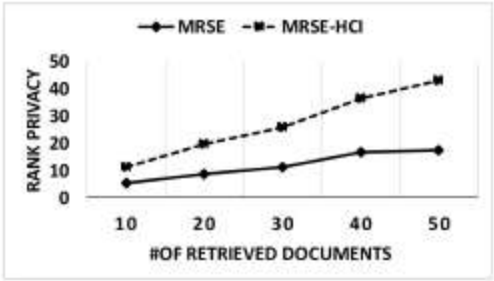


Fig: rank privacy

IV. CONCLUSION

In this script, we investigated ciphertext search inside the situation of cloud garage. We explore the trouble of preserving the semantic courting between extraordinary undeniable documents over the related encrypted files and provide the design method to beautify the overall performance of the semantic seek. We additionally advocate the MRSE-HCI structure to evolve to the requirements of information explosion, online data retrieval, and semantic search. At the identical time, a verifiable mechanism is also proposed to assure the correctness and completeness of seek effects. In addition, we analyze the hunt performance and safety under popular chance models. An experimental platform is constructed to assess the hunt performance, accuracy, and rank security. The experiment end result proves that the proposed architecture now not simplest properly solves the multi-key-word ranked search hassle, however additionally brings an improvement in seek performance, rank safety, and the relevance of retrieved files.

V. REFERENCES

[1]. M. Naor, and K. Nissim, Certificate revocation and certificate update, IEEE J. Sel. Areas Commun., vol. 18, no. 4, pp. 561-570, Apr. 2000.
[2]. H. Pang, and K. Mouratidis, Authenticating the query results of text search engines, Proc. VLDB Endow., vol. 1, no. 1, pp. 126-137, Aug. 2008.
[3]. C. Chen, X. J. Zhu, P. S. Shen, and J. K. Hu, ”A Hierarchical Clustering Method For Big Data Oriented Ciphertext Search,” presented at Proc. BigSecurity, Toronto, Canada, Apr. 27-May. 2, 2014.
[4]. S. C. Yu, C. Wang, K. Ren, and W. J. Lou, ”Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing,” in Proc. IEEE INFOCOM, San Diego, CA, 2010, pp. 1-9.
[5]. H. Witten, A. Moffat, and T. C. Bell, Managing gigabytes: compressing and indexing documents and images, 2nd ed., San Francisco: Morgan Kaufmann, 1999.
[6]. J. MacQueen, ”Some methods for classification and analysis of multivariate observations,” in Proc. Berkeley Symp. Math. Stat. Prob, California, USA, 1967, p. 14.
[7]. Z. X. Huang, Extensions to the k-means algorithm for clustering large data sets with categorical values, Data Min. Knowl. Discov., vol. 2, no. 3, pp. 283-304, Sep. 1998.
[8]. W. K. Wong, D. W. Cheung, B. Kao, and N. Mamoulis, ”Secure kNN Computation on Encrypted Databases,” in Proc. ACM SIGMOD, Providence, RI, 2009, pp. 139-152.
[9]. R. X. Li, Z. Y. Xu, W. S. Kang, K. C. Yow, and C. Z. Xu, Efficient multi-keyword ranked query over encrypted data in cloud computing, Futur. Gener. Comp. Syst., vol. 30, pp. 179- 190, Jan. 2014.
[10]. G. Craig. ”Fully homomorphic encryption using ideal lattices.” STOC. Vol. 9. 2009

Mr. Narahari Ajmeera is currently working as a Lecturer in Department of Computer Science and Engineering in Jawaharlal Nehru Technological University Hyderabad College of Engineering Manthani. I received M.Tech in Software Engineering from Kakatiya University and B.Tech Computer Science and Engineering from JNT University Hyderabad, Telangana, India.

Mrs. Sandhya Rani Pabbathi is currently working as Assistant Professor in Department of Computer Science and Engineering in Joginapally BR Engineering College. I received M.Tech in Computer Science from JNT University Hyderabad and BTech Computer Science and Engineering from JNT University Hyderabad, Telangana, India.

