ISSN 2278-3091
Volume 9, No.3, May - June 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse65932020.pdf
https://doi.org/10.30534/ijatcse/2020/65932020

Efficient Multi-User and Multi-Keyword Ranked Search Scheme for Encrypted Cloud Data

[1]D.Pradeepa ,[2]Dr.P.Sumathi
[1]Research Scholar, PG & Research Department of Computer Science, Government Arts College, Coimbatore INDIA.
pradeepamca09@gmail.com
[2]Assistant Professor, PG & Research Department of Computer Science, Government Arts College, Coimbatore INDIA.
sumathirajes@hotmail.com

ABSTRACT

With the improvement of cloud storage, greater data users are willing to outsource their data to cloud offerings. For privacy issues, sensitive data must be encrypted before outsourcing. There are numerous searchable encryption schemes to ensure statistics availability. However, the present search schemes pay little attention to the performance of facts users' queries, specifically for the multi-user situation. To allow the cloud servers to perform a search without understanding any statistics, to assemble a novel search and efficient technique based totally on the encrypted cloud statistics the usage of the “ECC cryptography scheme”. To obtain an efficient search, for every statistics person, a tree-based totally index encrypted with an additive order and secure characteristic is constructed. In order to rank the search outcomes, the proposed method utilize and model the relevance ratings of facts documents and advise a ``Iterative Deepening Depth-First Search" (ID-DFS) set of rules to attain the ranked results. To perform a keyword-primarily based query, the complete records set needs to be decrypted despite the fact that the matching end result set may be very small. It poses insufferable query latency and incurs unacceptable computational overhead. Finally, the proposed approach confirms the security and performance of the proposed scheme through complete theoretical evaluation and big experiments with real dataset.

Key words: - Cloud Services, Multi-keyword, Multi-User, encryption, decryption, cloud data.

1. INTRODUCTION

In latest years, good sized amounts of facts are produced via the usage of numerous assets such as tens of thousands and thousands of virtual sensors, social media packages, smart phones, financial transaction records, and so on. Thanks to many abilities furnished through cloud computing, statistics users and groups have appreciably moved their huge datasets from conventional local facts centres to the cloud that will make use of the opportunities of extra flexibility and decrease value. However, this calls for to be stored their sensitive statistics on remote untrusted servers and introduces new security and privacy disturbing situations that need to be treated. Therefore, the information is encrypted earlier than sending to the untrusted servers to be able to defend the facts confidentiality [1]. Although statistics encryption guarantees records confidentiality, it surely prevents the server from working at the records like keyword-based search over it.

However, various data owners encrypt their data with different keys leading to the following two drawbacks: [6] data users need to manage multiple keys for different data owners; [12] data users need to generate multiple trapdoors for data owners' data even for the same query condition. At the identical time, confidentiality of remotely stored statistics on untrusted cloud server is a huge trouble. In order to lessen the ones concerns, sensitive facts, which includes, personal health statistics, emails, earnings tax and financial critiques, are usually outsourced in encrypted shape the usage of famous cryptographic strategies. Although encrypted statistics storage protects far flung statistics from unauthorized get entry to, it complicates some simple, however critical records usage services together with plaintext key-phrase are looking for. A clean solution of downloading the records, decrypting and searching regionally is really inefficient given that storing information in the cloud is meaningless besides it could be easily searched and applied. Thus, clouds offerings ought to allow efficient search on encrypted records to offer the benefits of a super cloud computing environment.

Moreover, the data owner can share their data with a large number of users which requires the cloud server to have the ability to meet a large amount of requests with effective data retrieval services. One effective method for solving this problem is ranking the results and sending back the top-*K* files to the data user, rather than all of the relevant files [14]. This method can dramatically reduce the communication overhead and still meet user’s demand. However, such a ranking operation should not leak any other information related to the keywords.

Searching operations, and critical utility case while cloud-subsidized repositories boom in wide variety and duration, are proper examples where security, overall performance, and precision are relevant requirements. Yet present proposals for looking encrypted statistics are however restrained from more than one views, such as usability, query expressiveness, and purchaser-side overall performance and scalability [10]. This thesis focuses on the layout and assessment of mechanisms for searching encrypted statistics with superior overall performance, scalability, and usefulness and its miles involved with developing novel searchable encryption techniques that allow the cloud server to perform multi-key-phrase ranked are trying to find in addition to substring search

incorporating feature records. The proposed approach produce efficient solution for the problem statement given in the section 2 and the performance of the proposed approach is based on the searchable encryption schemes for ranked searching for and substring function search.

2. RELATED WORK

Cloud computing has been considered as a new model of corporation IT infrastructure, that can prepare big resource of computing, storage and packages, and allow users to experience ubiquitous, handy and on-demand community get entry to a shared pool of configurable computing sources with awesome efficiency and minimum financial overhead [3]. To assure statistics confidentiality, encryption is a great manner for users. But classical cryptographic primitives will cause a few important facts utilization offerings primarily based on plaintext to be inapplicable. In the cloud environment, statistics users typically share their outsourced files with other information users. Faced with the huge records, customers tend to search particular keyword to get their target documents. Encryption makes keyword search over encrypted cloud facts end up a brand new venture.

With the advent of cloud computing, facts users are influenced to outsource their complicated facts control systems from nearby websites to the economic public cloud for notable flexibility and economic financial savings. But for securing facts privacy, sensitive information should be encrypted earlier than outsourcing, which obsoletes traditional records utilization based totally on plaintext keyword search. Thus, permitting an encrypted cloud facts search provider is of paramount importance. Considering the massive wide variety of facts users and documents within the cloud, its miles necessary to permit more than one key phrases inside the search research and go back files inside the order of their relevance to these key phrases. Related works on searchable encryption recognition on unmarried keyword search or Boolean keyword searches, and seldom types the search consequences. The authors suggest first time; define and remedy the challenging hassle of privacy-preserving multi-keyword ranked search over encrypted statistics in cloud computing (MRSE). It establishes a set of strict privacy necessities for such a secure cloud statistics utilization system. Among numerous multi-keyword semantics, choose the efficient similarity degree of "coordinate matching," i.e., as many fits as viable, to capture the relevance of facts files to the search query and additionally similarly use "internal product similarity" to quantitatively examine such similarity measure. Here, first endorse a fundamental concept for the MRSE primarily based on comfortable internal product computation, after which deliver two drastically improved MRSE schemes to gain numerous stringent privacy necessities in two special hazard fashions. To improve search experience of the information search provider, suggest further amplify these two schemes to assist more search semantics [6].

Keyword-based totally search over encrypted outsourced facts has turn out to be an essential tool in the contemporary cloud computing state of affairs. The majority of the existing techniques are focusing on multi-keyword genuine healthy or unmarried keyword fuzzy search. However, those present strategies discover much less practical significance in real-world applications compared with the multi-keyword fuzzy search approach over encrypted data. The first try to assemble any such multi-keyword fuzzy search scheme turned into suggested through Wang et al., who used locality-sensitive hashing features and Bloom filtering to meet the purpose of multi-keyword fuzzy search. Nevertheless, Wang's scheme turned into simplest effective for a one letter mistake in keyword but turned into not effective for different common spelling mistakes. Moreover, Wang's scheme become liable to server out-of-order issues throughout the rating procedure and did now not recall the keyword weight. Here, based on Wang et al.'s scheme, proposed an efficient multi-keyword fuzzy ranked search scheme based on Wang et al.'s scheme this is capable of deal with the aforementioned problems. First, expand a new approach of keyword transformation primarily based at the uni-gram, to be able to simultaneously enhance the accuracy and creates the capacity to address different spelling errors. In addition, key phrases with the same root can be queried the use of the stemming algorithm [12].

With the growing recognition of cloud computing, massive amount of files are outsourced to the cloud for reduced control cost and simplicity of get right of entry to. Although encryption enables protective person statistics confidentiality, it leaves the properly-functioning yet practically-efficient at ease search functions over encrypted information a hard hassle. In the proposed technique present a verifiable privacy-preserving multi-keyword textual content search (MTS) scheme with similarity-based totally rating to cope with this trouble. To aid multi-keyword search and search end result rating, proposed to construct the search index based on time period frequency and the vector area version with cosine similarity degree to achieve higher search result accuracy. To improve the search efficiency, proposed a tree-primarily based index structure and numerous adaptive methods for multi-dimensional (MD) set of rules so that the sensible search performance is an awful lot higher than that of linear search. To in addition beautify the search privacy, and also proposed two secure index schemes to meet the stringent privacy requirements below strong hazard fashions, i.E., known ciphertext version and recognized historical past version [10].

The growing reputation of cloud computing, increasingly more information users are encouraged to outsource their information to cloud servers for tremendous comfort and decreased price in facts management. However, sensitive information must be encrypted before outsourcing for privacy necessities, which obsoletes statistics usage like keyword-primarily based document retrieval. The authors present a secure multi-keyword ranked search scheme over encrypted cloud information, which concurrently supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the extensively-used TF X IDF version are mixed in the index production and query generation. Here assemble a special tree-based index structure and suggest a "Greedy Depth-first Search" algorithm to offer efficient multi-keyword ranked search. The comfortable kNN algorithm is utilized to encrypt the index and query vectors, and in the meantime make sure accurate relevance rating calculation among encrypted index and query vectors [13].

With the arrival of cloud computing, it has become increasingly more popular for facts users to outsource their

facts to public cloud servers while allowing records users to retrieve this information. For privacy worries, secure searches over encrypted cloud statistics have motivated numerous research works beneath the single user model. However, most cloud servers in exercise do no longer just serve one user; alternatively, they aid a couple of users to proportion the advantages added via cloud computing. The authors proposed schemes to cope with Privacy preserving Ranked Multi-keyword Search in a Multi-user version (PRMSM). To allow cloud servers to carry out at ease search without understanding the real records of each keywords and trapdoors, and systematically construct a novel search protocol. To rank the search consequences and hold the privacy of relevance scores among key phrases and files, and proposed a novel Additive Order and Privacy Preserving Function family. To prevent the attackers from eavesdropping mystery keys and pretending to be criminal statistics users filing searches, and also proposed a singular dynamic secret key technology protocol and a new information person authentication protocol [9].

Cloud computing is becoming time-honored; information users are motivated to delegate complex facts managements to the industrial cloud for economic savings. Sensitive facts are usually encrypted before being uploaded to the cloud, which alas makes the frequently-used search characteristic a hard problem. The authors present a new multi-keyword dynamic search scheme with end result ranking to make search over encrypted information more at ease and realistic. In the scheme, employ a effective characteristic-hiding internal product encryption to decorate the security with the aid of preventing the leakage of search sample. For the priority of efficiency, the authors undertake a tree-based totally index structure to facilitate the searching process and updating operations [3].

Searchable encryption allows one to add encrypted files on a remote sincere-but-curious server and query that information at the server itself without requiring the documents to be decrypted previous to looking. The authors proposed a singular secure and efficient multi-keyword similarity searchable encryption (MKSim) that returns the matching facts objects in a ranked ordered way. Unlike all preceding schemes, this proposed search complexity is sub-linear to the whole range of documents that incorporate the queried set of keywords. The authors also analyse and demonstrate that proposed scheme is proved to be comfortable towards adaptive chosen-keyword attacks [5].

3. EXISTING APPROACH

It is the primary procedure to guard the data confidentiality is to encrypt the information earlier than outsourcing. Searchable Encryption is an emerging cryptographic process that lets in looking capabilities over encrypted records at the cloud. On this paper, a novel searchable encryption scheme for the customer-server shape has been offered. The scheme exploits the homes of the modular inverse to generate a probabilistic trapdoor which helps the search over the search inverted index table. On this procedure, the task of assisting key phrase search on encrypted data outsourced to the cloud. There are making of more than one contribution to this area through presenting a singular ranked targeted searchable encryption scheme [7]. The current technique constructs and exploits the houses of modulo high to generate a probabilistic trapdoor. The greatest challenge in searchable encryption is to preserve stability between protection, affectivity and query expressiveness.

The present method was indistinguishability that is executed with the useful resource of making use of the belongings of a probabilistic trapdoor. It is designed and carried out a proof of perception prototype and checks the proposed scheme with a real dataset of files [7]. It analyzes the performance of scheme in competition to the claim of the scheme being mild weight. The safety assessment yields that this scheme assures a better diploma of protection compared to other latest schemes.

3.1. Difficulties in Existing Approach

- It handiest helps the single keyword ranked search encryption scheme.
- Huge fee in phrases of information usability. For instance, the existing techniques on key-phrase-based data retrieval, which can be appreciably used on the plaintext records, can't be immediately carried out on the encrypted facts. Downloading all the statistics from the cloud and decrypt locally is obviously impractical.
- Existing System strategies not practical because of their excessive computational overhead for each the cloud server and client.

4. PROBLEM STATEMENT

Cloud storage permits ubiquitous, scalable, and on-call for community access to a shared pool of virtual information assets. More enterprises and people generally tend to outsource their personal data to the cloud server, and make use of query offerings to without difficulty access information anytime, anywhere and on any tool.

In order to fulfill the sensible search necessities, search over encrypted records should help the subsequent three capabilities. First, the searchable encryption schemes need to aid multi-keyword search, and provide the equal user experience as looking in Google search with exceptional keywords; single-keyword search is some distance from quality by using handiest returning very constrained and erroneous search consequences [11]. Second, to quick perceive most relevant consequences, the search user could typically opt for cloud servers to kind the back search outcomes in a relevance-primarily based order ranked via the relevance of the hunt research to the documents. In addition, displaying the ranked search to customers also can get rid of the needless network traffic by means of handiest sending back the most relevant effects from cloud to look users. Third, as for the search performance, for the reason that number of the files contained in a database will be noticeably huge, searchable encryption schemes have to be efficient to quickly respond to the search researchers with minimal delays [9].

4.1. Design Goals of the Proposed Approach: -

To enable an efficient ranked multi-keyword search for a couple of user over encrypted cloud information, the proposed scheme targets to reap the subsequent desires [8]:

- Multi-keyword Ranked Search for Multiple Users.

- Search Efficiency.
- Security.

5. PROPOSED APPROACH

The Proposed Methodology is a comfortable tree-based completely are searching for scheme over the encrypted cloud statistics, which supports multi-key-phrase ranked are searching for and dynamic operation on the file collection[2]. Specifically, the vector area model and the widely-used "time period frequency (TF) × inverse file frequency (IDF)" version are blended within the index creation and query technology to offer multi-keyword ranked are trying to find. In order to attain excessive are searching for performance, the proposed approach process a tree-based index shape and suggest a "Iterative deepening depth-first search (ID-DFS)" algorithm primarily based in this index tree. The secure Support Vector Machine (SVM) set of rules is applied to encrypt the index and query vectors, and in the intervening time make certain correct relevance score calculation among encrypted index and query vectors. To resist unique attacks in distinct danger models, the enhanced dynamic multi-keyword ranked search scheme would be used [5]. The following figure 1, shows the framework of the proposed approach. It consists of four modules.

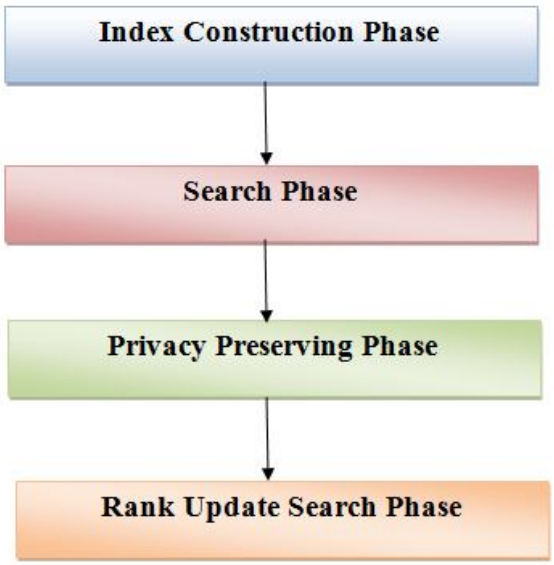

Figure 1: Proposed Approach Framework

4.1. Index Construction Phase

This Phase allows the user to register the ones details and also embody login info. This Phase helps the user to feature his document with encryption the usage of ECC algorithm set of rules. This guarantees the files to be included from unauthorized man or woman. Data user has a set of documents F =f_1; f_2; :::; f_n that he desires to outsource to the cloud server in encrypted form on the equal time as nonetheless preserving the capability to search around on them for effective utilization. In the proposed scheme, the information user first of all builds a duplicate searchable tree index I from record series F, after which generates an encrypted report series C for F. Afterwards, the records user outsources the encrypted series C and the search index I to the cloud server, and securely distributes the important thing data of trapdoor technology and record decryption to the authorized data customers. Besides, the statistics user is responsible for the update operation of his files stored inside the cloud server. While updating, the facts user generates the update facts domestically and sends it to the server.

4.2. Search Phase

This Phase is used to help the purchaser to go looking the record the use of the couple of key terms concept and get the accurate result list based totally on the user query [3]. The character goes to pick out the specified document and sign up the user data and get activation code in mail e mail earlier than enter the activation code. After user can download the Zip file and extract that report. Data customers are criminal ones to get admission to the files of records user. With t query key terms, the felony user can generate a trapdoor TD in preserving with search manage mechanisms to fetch adequate encrypted files from cloud server. Then, the statistics user can decrypt the files with the shared secret key.

4.3. Privacy Preserving Phase:

This Phase is used to help the server to encrypt the report using Elliptic-curve cryptography (ECC) Algorithm and to convert the encrypted record to the Zip document with activation code and then activation code deliver to the person for down load. Cloud server shops the encrypted document collection C and the encrypted searchable tree index I for statistics user. Upon receiving the trapdoor TD from the records user, the cloud server executes search over the index tree I, and ultimately returns the corresponding series of top-k ranked encrypted documents. Besides, upon receiving the update statistics from the statistics user, the server wants to replace the index I and document collection C consistent with the obtained statistics.

4.4. Rank Update Search Phase

This Phase make sure the person to go looking the files which can be searched often the usages of rank are trying to find. This Phase lets in the client to down load the document the usage of his secret key to decrypt the downloaded statistics. This Phase lets in the User to view the uploaded documents and downloaded files. The proposed scheme is designed to provide no longer satisfactory multi-keyword query and accurate cease result rating, but also dynamic update on report collections. The scheme is designed to prevent the cloud server from mastering extra information approximately the document collection, the index tree, and the query.

5. PERFORMANCE EVALUATION METRICS

We implement the proposed scheme using Python on a Windows 7 operation system with Intel core i5 processor 3.30 GHz. Here leverage Pairing-Based Library to simulate the cost of cryptographic operations. The document set we test on is the collection of Request for Comments (RFC). The keyword dictionary is extracted from document collection according to traditional criterion [4]. The relation of document number *n* and dictionary size *m* can be used. The experiments consist of the performance of index construction, search and update processes.

In this segment, the proposed approach is evaluated an intensive experimental assessment of the proposed method on real records set: The overall performance of the proposed approach is evaluated concerning the efficiency of the proposed scheme, in addition to the tradeoff between search precision and privacy[4]. In this section the designated overall performance of the proposed machine is analyzed using the following metrics.

Precision:

Precision is the ratio of effectively predicted wonderful observations to the whole anticipated high-quality observations**.**

Accuracy:

Accuracy is the most intuitive overall performance measure and it is virtually a ratio of correctly predicted statement to the entire observations. One may think that, if the excessive accuracy then the proposed model is quality**.**

Recall:

Recall is the ratio of efficiently anticipated positive observations to the all observations in real elegance**.**

F-measure:

It may be a higher unmarried metric when compared to precision and recall; each precision and recall give unique records that could complement each other when blended.

6. CONCLUSION

In this paper, for the first time the proposed system outline and clear up the trouble of multi-user and multi-keyword ranked find over encrypted cloud statistics, and establish a diffusion of privacy requirements. Among numerous multi-keyword semantics, here, select the secure and efficient similarity measure as many suits as viable; to successfully seize the relevance of outsourced documents to the query key phrases, and use the similarity to quantitatively compare such similarity degree. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world records set show the proposed schemes introduce low overhead on both computation and communication.

REFERENCES

[1] Li, H., Liu, D., Dai, Y., Luan, T. H., & Shen, X. S. (2014). Enabling efficient multi-keyword ranked search over encrypted mobile cloud data through blind storage. *IEEE Transactions on Emerging Topics in Computing*, *3*(1), 127-138. https://doi.org/10.1109/TETC.2014.2371239

[2] Li, J., & Chen, X. (2013). Efficient multi-user keyword search over encrypted data in cloud computing. *Computing and Informatics*, *32*(4), 723-738.

[3] Yan, J., Zhang, Y., & Liu, X. (2016). Secure multi-keyword search supporting dynamic update and ranked retrieval. *China Communications*, *13*(10), 209-221. https://doi.org/10.1109/CC.2016.7733045

[4] Nabil, M., Alsharif, A., Sherif, A., Mahmoud, M., & Younis, M. (2018, May). Efficient multi-keyword ranked search over encrypted data for multi-data-owner settings. In *2018 IEEE* International Conference on Communications (ICC) (pp. 1-6). IEEE. https://doi.org/10.1109/ICC.2018.8422281

[5] Strizhov, M., & Ray, I. (2014, June). Multi-keyword similarity search over encrypted cloud data. In IFIP international information security conference (pp. 52-65). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-55415-5_5

[6] Cao, N., Wang, C., Li, M., Ren, K., & Lou, W. (2013). Privacy-preserving multi-keyword ranked search over encrypted cloud data. IEEE Transactions on parallel and distributed systems, 25(1), 222-233.

[7] Tahir, S., Ruj, S., Rahulamathavan, Y., Rajarajan, M., & Glackin, C. (2017). A new secure and lightweight searchable encryption scheme over encrypted cloud data. *IEEE* Transactions on Emerging Topics in Computing.

[8] Peng, T., Lin, Y., Yao, X., & Zhang, W. (2018). An efficient ranked multi-keyword search for multiple data owners over encrypted cloud data. IEEE Access, 6, 21924-21933.

[9] Zhang, W., Lin, Y., Xiao, S., Wu, J., & Zhou, S. (2015). Privacy preserving ranked multi-keyword search for multiple data owners in cloud computing. IEEE Transactions on Computers, 65(5), 1566-1577. https://doi.org/10.1109/TC.2015.2448099

[10] Sun, W., Wang, B., Cao, N., Li, M., Lou, W., Hou, Y. T., & Li, H. (2013, May). 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 (pp. 71-82). https://doi.org/10.1145/2484313.2484322

[11] Yang, Y., Liu, X., & Deng, R. (2017). Multi-user multi-keyword rank search over encrypted data in arbitrary language. *IEEE* Transactions on Dependable and Secure Computing.

[12] Fu, Z., Wu, X., Guan, C., Sun, X., & Ren, K. (2016). Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement. IEEE Transactions on Information Forensics and Security, 11(12), 2706-2716. https://doi.org/10.1109/TIFS.2016.2596138

[13] Xia, Z., Wang, X., Sun, X., & Wang, Q. (2015). A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data. *IEEE transactions on parallel and distributed systems*, *27*(2), 340-352. https://doi.org/10.1109/TPDS.2015.2401003

[14] Xiangyang, Z., Hua, D., Xun, Y., Geng, Y., & Xiao, L. (2017). MUSE: an efficient and accurate verifiable privacy-preserving multikeyword text search over encrypted cloud data. Security and Communication Networks, 2017. https://doi.org/10.1155/2017/1923476