Research Article

A Survey on Data Retrieval Techniques Over Encrypted Cloud Storage

Arpitha T V,[1*] Mallikarjuna Shastry[1]

Abstract: Improvement in Cloud computing has changed the perspective of data owner which propel the information technology to outsource their information to people in general cloud server like Amazon, Microsoft Azure, Google Drive, and so forth. With the assistance of cloud computing, the clients can get to information outsourced by data owner remotely from anyplace whenever and pay just for what they are utilizing. CSPs (Cloud Service Provider) deal with its information and its protection however there are a percentage of the variables in view of which the information security and client personality might be disregarded. Along these lines, information proprietors ought to encode their separate touchy information before outsourcing it to the general population cloud server. The information is getting encrypted before outsourcing which may influence the execution of some vital information getting to operations like looking of an archive, and so on. Various arrangements have been offered to make the recovery of information from the CSP in an effective and secure way. CSPs assumes an imperative part for information security, however is it adequate for the touchy information like well being record, pictures, secret key, finance time sheets, and so forth? In this way, to answer the inquiry, there are some best techniques/arrangement offered to give security and protection to the information over cloud server. In this review few of the keyword searching techniques has been examined, to find a successful techniques/answer for the retrieval of information/documents over encrypted cloud information.

Asian Journal of Engineering and Technology Innovation
Volume 4, Issue 7
Published on: 7/05/2016

Cite this article as: Arpitha T V, Mallikarjuna Shastry. A Survey on Data Retrieval Techniques Over Encrypted Cloud Storage. Asian Journal of Engineering and Technology Innovation, Vol 4(7): 43-48, 2016.

INTRODUCTION

Today more services like documentation, storage, work processes, email and office applications like book keeping, HR, buy, CRM among others are being conveyed structure the cloud. There are three large acknowledged cloud service models - Software as a Service (SaaS), Infrastructure as a Service (IaaS) and Platform as a Service (PaaS). New service models can be gotten from these three essential cloud service conveyance models and as a particular cloud service model. Cloud Computing offers answers for some issues like hardware, machine failures and so on. The immense point of interest of cloud computing is versatility the capacity to include limit or applications nearly immediately. The pay-as-you-go approach engages little and medium-sized undertakings; since the seller has numerous clients, it can bring down the per-unit expense to every client. Bigger organizations think that its less demanding to oversee coordinated efforts in the cloud.

Data Retrieval includes separating the needed data from a database. To recover the coveted data the client show an arrangement of criteria by a query. At that point the Data base management framework, programming for overseeing databases, chooses the required data from the databases. To maintain a strategic distance from any sorts of assault, the touchy data ought to be encrypted, before it is outsourced which obsoletes customary data use in light of plaintext keyword search. Subsequently, empowering an encrypted cloud data search administration is of fundamental significance. Considering the expansive number of data clients and records in cloud, it is essential for the search administration to permit keyword query and give result closeness positioning to meet the powerful data recovery required.

RELATED WORKS

Studied varies encrypted data retrieval technique likes Secured Multiple-Keyword Search (SMS)[1], It helps to develop safe cloud data consumption method. [2] To prevent leakage of data uses Order Preserving Encryption (OPE) by utilizing multi-keyword search. [3] Describes the privacy and removing of data leakage using Two Round Searchable Encryption (TRSE). [4] Helps to secure delicate data by improving index structure. [5]Uses Multi-Keyword Text/keyword Search (MTS) for secure cloud search capacities over encrypted data. [6] Describes on realizing secure semantic search through query keyword semantic

[1]Reva Institute of Technology and Management, Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Near Border Security Bustop, Bengaluru, Karnataka-560064, India.
E-mail: arpitha.tv@gmail.com
*Corresponding author

extension technique based on the co-occurrence probability of terms.

DATA RETRIEVAL TECHNIQUE

Secured Multiple Keyword Search over Encrypted Cloud Data

C.R. Barde, PoojaKatkade, DeepaliShewale, RohitKhatale et al describes the problem of Secured Multiple-keyword Search (SMS) on encrypted cloud data. It constructs a group of confidentiality policies for safe cloud data consumption method [1] .

This reference paper has picked a standard called coordinate matching, which is utilized to distinguish the similitude between search inquiry and data documents. This paper additionally pick client's inward data correspondence which is utilized to compute the similarity of one archive with the search inquiry in matching the coordinates technique. Each record is associated alongside a vector as a sub-index and each piece shows whether proportionate keyword is available in the report or not. Thus, to accomplish multi-keyword searching with index and search confidentiality, a fundamental SMS plan with internal item count is utilized.

The objectives are:

1. Resolving Multi-keyword search problem on cloud data.
2. Presenting two patterns which track the standard of inner product similarity and coordinate matching.

Mentioned the existing searchable encryption scheme which allows clients to firmly search above encrypted cloud data, using keywords but won't go for decryption. The drawbacks of this system are: Single keyword search without ranking, Boolean keyword search without ranking, Failed to get relevant data.

Because of the disadvantages, [1] proposed a system which follow the method of coordinate matching, to recognize the similarity between data documents and search query. It is used to calculate the resemblance of a document to with the search query in coordinate matching standard.

SMS system uses secure inner product calculation is occupied starting from a secure k-Nearest Neighbor (kNN) method. It increases various privacy needs in 2phases of threat models.

1. Showing the difficulty of Multi-keyword search system on encrypted cloud data
2. Suggest two patterns following the principle of inner product similarity and coordinate matching.

It consists of 3 modules:

1. Encryption Module: This module underpins Data Owner. Data Owner encodes the document by RSA Algorithm later it sends the record to a zip format a long with a generated activation code. At that point the same activation code will be given to the client for downloading.

2. Multi-keyword Module: This module underpins User, in getting the precise results taking into account the idea of multiple keywords. Users can enter more than a word in their question; the server is going to join every one of the inquiries into an unmistakable single word. At long last, server gives the equal rundown of words from the database lastly the client gets the document from the rundown.

3. File Upload Module: This module bolsters cloud server to watch record transferring in a secured way. Administrator will login utilizing the log keys, before the administrator logout it is workable for administrator to adjust the log keys. The administrator can likewise change the secret key after each login and can see the client downloading history. They can likewise see the record demand tally information on stream diagram. The administrator needs to transfer the document strictly when the transformation of Zip record format.

Algorithms Used

1. RSA Algorithm: This algorithm is used for the purpose of encrypting and decrypting the data files. It is an asymmetric algorithm, as it uses two dissimilar keys for encryption and decryption.

RSA algorithm contains three different steps: Generation of Key, Encryption and Decryption.

Generation of Key: RSA contains a private key and a public key. Public key is identified by every user and it is can be applicable for encrypting data sets. Data encrypted with a public key can be decrypted using a private key.

2. K-Nearest Neighbor: K-nearest neighbor search recognizes the top k nearest neighbors to the query. This technique is used in predictive analytics to calculate a point based on the consensus of its neighbors. K-nearest neighbor graphs are the graphs in which every point is linked to its k nearest neighbors.

The significance of dmax is reduced with the on-going particular estimation of the object resemblance distance for the users. At the end of the step by step modification, dmax influences the optimal query range Ed. This avoids the system from creating more users than essential and satisfies the r optimality principle.

Worked on the trouble of multiple-keyword search over cloud data alongside the ranking scheme, it builds multiple security parameters. To do that it considered a proficient rule of coordinate matching. It builds secure inward data calculation; furthermore it accomplished viable ranking result utilizing k-Nearest Neighbor technique [1]. This framework worked just over specific single cloud, and in forthcoming days it will be handled up to cloud computing to offer enhanced security in multi client strategies.

Retrieval of Encrypted Cloud Data Using Multi-Keyword

C. Rajeshkumar, Dr.K.RubaSoundar et al portrays the idea of Encrypted Cloud data recovery utilizing multi-keyword. As and when the users send their own data onto the cloud, the administration supplier ought to be equipped for dealing with the data and the link between the users and cloud. The issue of data confidentiality is one of the significant issues in the cloud. With the end goal of data confidentiality, delicate data must be encrypted some time ago sending it to cloud [2].

Ranking on server side is set up utilizing Order Preserving Encryption (OPE) which can possibly bring about leakage of data. With OPE, it is conceivable to do Boolean search. Boolean search is a type of searching technique; these licenses users to blend one or more keywords utilizing administrators like AND, NOT as well as to make further correct matching results

Searchable Symmetric Encryption (SSE) is utilized with the end goal of secure data retrieval from cloud. SSE is utilized to get the topmost related documents that equivalent client's necessity in the event of every single record. This orders the documents must be kept in ranked request of criticalness in view of the significance of users and the records with most extreme centrality's are given to users. To maintain a strategic distance from data leakage, SSE proposes a Two Round Searchable Encryption (TRSE), it gives top-k multi-keyword retrieval. TRSE utilizes Vector Space Model (VSM) and homomorphic encryption, which takes out the information leakage and permits secure capacity of data in cloud.

The main objectives are:

1. To store the Encrypted data in cloud
2. Retrieving the Encrypted Cloud Data Using Multi-Keyword.

Searchable Symmetric Encryption (SSE) is utilized with the end goal of secure data retrieval from cloud. SSE is utilized to get the topmost related records that equivalent client's necessity if there should arise an occurrence of every single document. This orders the documents must be kept in ranked request of criticalness in light of the significance of users and the records with most extreme hugeness' are given to users. To keep away from data leakage, SSE proposes a Two Round Searchable Encryption (TRSE), it gives top-k multi-keyword retrieval. TRSE utilizes Vector Space Model (VSM) and homomorphic encryption, which disposes of the information leakage and permits secure capacity of data in cloud.

It consists of five main modules:

1. Setup Phase: In Setup phase the Owner of the data (Data Owner) produces a public key and a secret key for the homomorphic encryption system. Homomorphic encryption is a kind of encryption which allows particular types of computation that can be performed on cipher text to gain plain text.
2. Index Build Phase: Data Owner constructs the searchable index using the set of files called C. Techniques from Information Retrieval like stemming is used to construct searchable index from C. Then it is encrypted with private key and outputs the secure searchable index.
3. Trapdoor Generation Phase: Data user creates secure trapdoor using the request. Vector is constructed using user's multi- keyword appeal later it is encrypted as secure trapdoor with public key from private key, outputs the secure trapdoor called T.
4. Score Calculation Phase: When cloud server receives secure trapdoor, it calculates the total scores of every file in index and proceeds with the encrypted gives vector to the data user.
5. Rank Phase: Here data user secret key to decrypt the vector and then raise an appeal to get files with top-k scores.

To explain the problem of multi-keyword top-k accessing on encrypted cloud data, motivate a solution by describing similarity relevance and scheme robustness[2] .

Top-K Retrieval of Encrypted Cloud Data Using Secure Multi-Keyword

K. Manoj Kumar, E. Purushotham et al describes the cloud data privacy and elimination of information leakage in cloud using Two Round Searchable Encryption (TRSE) [3].

In cloud computing environment data owner keeps up an arrangement of documents to send its encrypted structure into the cloud server. Towards this, the data owner needs to make a searchable index utilizing the keywords and afterward it sends both the encrypted index and encrypted records into the cloud server. The approved data client first makes a question demand and the cloud server sends coordinated documents to the data client. To dispose of the information leakage, TRSE scheme has been utilized, that procurements top-k multi-keyword retrieval. Homomorphic encryption and Vector space model are utilized that include as a part of ranking. Since ranking is done on client side in light of request safeguarding encryption (OPE) proficiency in retrieval of document is moved forward. The documents are requested in the importance by client's notification and the records with highest essentialness are sent to users.

Keyword-based retrieval is a great information framework and broadly utilized as a part of plaintext circumstances, client's entrance fitting records in bolstered words. A grouping of searchable general encoding (SSE) schemes is utilized to modify search over figure text. To upgrade security while not losing power, techniques are given which demonstrates how they bolster top-k single keyword retrieval beneath various consequences. The creators make an endeavor to determine the matter of top-k multi - keyword over encrypted cloud information. Creators have exhibited the thoughts of comparability association and topic vigor to express the protection related issues in searchable schemes, and after that it illuminates the vulnerability drawback by proposing two round searchable encryption (TRSE) technique [3].

1. TRSE Design

A percentage of the present schemes utilizes server-side requested ranking bolstered OPE to upgrade the strategy of data retrieval over encrypted cloud knowledge. Yet, server side rank backings OPE that upsets the protection of delicate information that is taken into record uncompromisable inside the security concentrating outsider cloud registering situation. To acknowledge knowledge confidentiality, ranking is left to the client.

2. Framework of TRSE

The framework of TRSE includes four algorithms: Setup, index build, Trapdoor Gen and Score Calculate;

3. Setup

The info owner produces a key and public keys to do homomorphic encoding. The protection factor lambda is used, because the input and the output can be a secret key, and a public key set.

4. Relevance Rating

Some of the multi key word SSE schemes support matches a question. Considering the big range of information clients and files within the cloud, it's essential to allow Multi key word within the search question and come back files within the rank of their connection with the inquired keywords.

5. Vector Area Model

While TF-IDF represents the burden of one keyword on a document, it tends to use the vector area model to attain a file on multikey word. The vector area model is a pure mathematics entity for demonstrating a file as a vector.

6. Homomorphic Encoding

Homomorphic encoding could be a kind of encoding which allows particular varieties of calculations to be passed out on cipher text and procure an encrypted outcome that decrypted equals the results of actions achieved on the plaintext.

Discusses the importance of TRSE system, by comparing its simplicity. This easiness is attained at the price of an oversized key size [3].

Indexing and Advanced Relevance Ranking Score Preserving for Multi-Keyword Search over Encrypted Cloud Data

Amol D. Sawant, Prof. M.D Ingle et al depicts the issue of saving total of multiple keywords search records ranking score [4]. This additionally enhances the ranking score of the documents which expands the framework ease of use by giving critical requested ranking as opposed to giving indistinguishable yield. It finds the measurable degree strategy i.e. enhanced centrality score from the information retrieval to make the index of records and to enhance the propelled ranking capacity. The progressed searching is utilized to secure the entirety of the delicate score information by enhancing the index structure. Confidentiality related issues to get to the data from cloud contain two strategies:

1. Sensitivity of keywords utilized as a part of the search demands
2. Sensitivity of related data got to by the client

The main objective of the approach is:

1. To achieve efficient searching using multi-keyword technique
2. To achieve relevance based ranking using synonym technique.

The scheme addressed can do multiple keyword searches in a solitary question and gives the ranked results to the client [4]. Client can get to just the most noteworthy matches and thus the season of users will be spared to get to the essential data. In the event that client needs to utilize the multiple keyword search then the total of the significance score of the records should be figured to get to the documents. The undertaking is fitting in keeping the total of the significance score of the multiple keyword documents. The scheme builds the index for multi-keyword search; it additionally diminishes the quantity of particular keywords and the utilization of multilevel indexing to spare the searching time and to give the propelled ranking consequences of the report.

The issue of keyword search on cloud server is that the server may know the search terms and the protection of the keyword won't be secure. The searching ought to be done in encrypted format and the data is likewise put away in encrypted format. The encryption keys ought not to know to the database server. There are three modules in this system architecture are:

1. Data Owner, genuine database owner who makes the database and client is unequipped for taking care of the database, so client stores the data in encrypted format in the cloud server..
2. Users, who is approved from the Data Owner, client utilizes the administrations and data conveyed by the Data Owner through cloud server. Users search for the data on cloud server utilizing the encrypted keywords and access the cloud in the wake of acquiring the authorization from the Data Owner.
3. Cloud Server, is the fundamental cloud database server and keeps up the enormous measure of encrypted data and handles the quantity of solicitations from number of users and offer the reliable support of the client and complete the encrypted searching of the archive.

The steps included in Index Generation stage and Trapdoor Generation stage are:

Step 1: Data Owner encodes the built index and transfers it to cloud server.

Step 2: Data Owner transfers the encrypted records into cloud server.
Step 3: User solicitations to execute a searching operation on keywords, thus client first contacts the Data Owner and solicitations for the trapdoor
Step 4: After acquiring the trapdoor, client creates an encrypted inquiry and solicitations the cloud server for the data.
Step 5: The cloud servers perform encrypted search and ranks the relating reports and gives encrypted result to the client .
Step 6: User ask for the decoding keys from the Data Owner and acquire the keys for the documents and unscramble the required records.

The objective is to shroud the entirety of the importance score of the multiple keywords in multi-keyword ranked search. To do the multi-keyword search, the whole of the pertinence score of keywords of documents is preserved from the server. To do the ranking of the multiple keywords the server need to compute the total of the significance score of each keyword, then makes a correlation between the got whole and server will figure the ranking of records to appear to users. In the event that the server knows the pertinence score then it will distinguish the configuration and the outcome will be leaked to the server.

Likewise enhance the pertinence ranking capacity to enhance the aggregate score of the record. The ranked search work fundamentally enhances the essentialness of the returned result and it diminishes the correspondence overhead, which is amazingly required [4]. The conjunctive ordinary search is executed to actualize the multi-keyword search. The test outcome will deliver the lessening in the particular keywords utilizing propelled IR techniques, additionally enhances the pertinence ranking score of the documents. As we are utilizing the multi-keyword search, search requires additional time than single keyword search, here future upgrade can be to lessen the search time.

Privacy-Preserving Multi-Keyword Search in Cloud supporting Similarity based Ranking

Wenhai Sun, Bing Wang,Ning Cao, Ming Li,Wenjing Lou, Y. Thomas, Hui Liet al depicts the issue of client data confidentiality and secure cloud search capacities over encrypted cloud data [5]. They tended to the issue utilizing security defensive Multi-Keyword Text/Keyword search (MTS) technique alongside equivalent word ranking.

It proposes how search index will be established by Term Frequency (TF) and the Vector Space Model (VSM) alongside level of cosine comparability. It can be utilized as a part of request to achieve propelled search yield precision. With a specific end goal to expand search fitness, recommends index structure and Multi-Dimensional (MD) algorithm [5]. The principal arrangement for multi-essential word content request with likeness based situating is secure under known figure content model.
The responsibilities of [5] are dense as takes after:
1. By melding the best in class data recuperation methodology, propose a security ensuring multi-catchphrase content interest arrangement supporting likeness based situating.
2. It proposes a randomization strategy in the enhanced arrangement to dodge delicate repeat data spillage thusly fulfilling better security of critical words.
3. Enhanced secure index scheme is still practically identical in search time to Multi-keyword text search along through likeness based ranking.

1. Accuracy-improved Multi-keyword Ranked Search

To plan a scrambled cloud information hunt plan which is not just backings the viable multi-catchphrase seek usefulness, additionally, by selection of the vector space model, accomplishes the exactness enhanced closeness based item positioning?

2. Search Efficiency

In the place of linear search it discovers a tree-based index structure and an effective search algorithm to attain improved applied search competence.

3. Privacy Goals

The main aim is to safeguard client's confidentiality by avoiding the cloud server from knowing info of the files, the index tree, and the queries.

Utilizes connected and agent multi-keyword data search strategy as an early push to determine the issue of secure searching on encrypted cloud. The two noteworthy components are, supporting likeness based ranking for further right search result and a tree-based search algorithm that achieves enhanced direct search profitability.

Achieving Efficiency of Encrypted Cloud Data with Synonym Based Search and Multi-Keyword Ranked Search

DipikaChavan, Dinesh Yadav et al depicts on acknowledging secure semantic search through question keyword semantic expansion technique taking into account the co-event likelihood of terms[6] .

To accomplish proficiency of searching technique, utilizes Term Frequency-Inverse Document Frequency (TF-IDF) algorithm with keyword set and its semantic words. This backings the retrieval of data utilizing semantic inquiry. To achieve effective searching on encrypted cloud data, its focused on an efficient and flexible searchable scheme which supports both multi-keyword ranked searching and similarity based search.

The level of cosine is connected to compute the likeness between the document and the search question. To expand the viability of the search technique use amplified keyword set with

semantic words or natural language words for the keywords. This backings data access on semantic question. Regardless of the fact that client doesn't know accurate or synonym of keywords of encrypted data, client can at present do searching by its significance in natural language.

1. To accomplish proficiency in searching over encrypted cloud data without co-working the data confidentiality
2. Solving the issue of proficient searching utilizing closeness based multi-keyword technique

The current techniques like genuine or fuzzy keyword searching frameworks are pertinent for just single keyword searching. These schemes doesn't get to the suitable data to users inquiry henceforth multi-keyword ranked search over encrypted cloud data is presented.

While client searching the data on cloud server, the client may have no knowledge of the accurate words to do searching. To increment searching, it is essential for the search motors to comprehend what the client needs, and after that no one but they can give the outcome in legitimate request. To accomplish this, one of the crucial things that builds the proficiency is semantic based searching. By joining metadata highlights, semantic searches can be fundamentally created when contrasted with out-dated search techniques like fuzzy and definite matching. Semantic search permits users to utilize natural language to express what client needs to find. The E-TFIDF algorithm is utilized to enhance the archives searching improvement, for exact circumstances where client need to discover a record however don't have entry the genuine words utilized.

The algorithm used in this reference paper [6] presents:

1. The number of direct words of the searching present in the document.
2. The number of word difference of the searching present in the document.

The number of similar words of the keyword in the searching present in the document

CONCLUSION

From the above review, it is clear that searching on encrypted cloud data has part of difficulties numerous searchable techniques have been examined in view of single keyword, multiple keyword search, Ranking, Similarity search, Fuzzy resilience. A definitive objective is to empower rich search semantics in a security saving way and productively bolster for vast scale and disseminated nature of cloud data. Past the text data search, numerous methods for speaking to the data are accessible for instance diagram organized data search, picture data search, multimedia data search and multi dimensional search. This concentrate just explained different data retrieval techniques for text data.

REFERENCES AND NOTES

1. C. R. Barde, Pooja Katkade, DeepaliShewale and RohitKhatale, “Secured Multiple-keyword Search over Encrypted Cloud Data ”,International Journal of Emerging Technology and Advanced Engineering , Volume 4, Issue 2, February 2014
2. C.Rajeshkumar and Dr.K.Rubasoundar, “Retrieval of Encrypted cloud data using multi-keyword”, International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Special Issue 1, March 2014.
3. K. Manoj Kumar, E. Purushotham, “Top-K Retrieval of Encrypted Cloud Databy Using Secure Multi-Keyword”, International Journal of Software and Hardware Research in Engineering, Vol 2,Issue 8, August 2014
4. Amol D. Sawantand Prof. M.D Ingle,” Indexing and Advanced Relevance Ranking Score Preserving for Multi-Keyword Search over Encrypted Cloud Data”, International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (3),3165 – 3169,2014
5. Wenhai Sun, Bing Wang, Ning Cao, Ming Li,Wenjing Lou, Y. Thomas Houand Hui Li, “Privacy-preserving Multi-keyword Text Search in the CloudSupporting Similarity-based Ranking”, ASIA CCS, 2013
6. DipikaChavan, Dinesh Yadav et al [10] describes on realizing secure semantic search through query keyword semantic extension technique based on the co-occurrence probability of terms.