International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 01 | Jan 2020 www.irjet.net p-ISSN: 2395-0072

An Efficient Solitude Securing Ranked Keyword Search Technique

Asha Elsa George[1], Bibin Varghese[2], Smita C Thomas[3]

[1]P G Scholar, Dept. Of CSE, Mount Zion College of Engineering, Kadammanitta, Kerala, India
[2]Assistant Professor, Dept. Of CSE, Mount Zion College of Engineering, Kadammanitta, Kerala, India
[3]Research Scholar Vels University, Chennai, India

Abstract - As the volume of information in server farm is encountering a huge development, so proprietor of information want to re-appropriate delicate and significant reports with the end goal of security monitoring. The archives are put away in scrambled configuration so it is fundamental to create productive hunt figure content procedure. During the time spent encryption connection between records is hidden which prompts perform crumbling. A quality various leveled bunching (QHC) strategy is proposed to help looking through component and to meet quick looking inside cloud condition. Multi catchphrase positioned search various leveled grouping file (MRSE-HCI) engineering is utilized. For query output check, least hash sub tree is utilized. The proposed strategy has a few favorable circumstances over customary technique in archive's recovery and rank protection.

Key Words: Cloud Computing, Hierarchical Clustering, Security, Cipher text search, Multi keyword search, Ranked search.

1. INTRODUCTION

As we venture into advanced stage, terabyte's of information is created overall every day, so any information proprietor who need to redistribute their information should be venture up in a distributed computing. Associations and enterprisers with huge measure of information like to redistribute information so as to lessen information the executives cost and storeroom. In spite of the fact that cloud specialist organization's (CSP) guarantee their administrations in regards to security, protection measure security and protection are significant deterrents forestalling for more extensive acknowledgments.

A traditional method for information encryption makes server side applications, for example, looking on scrambled information turns into danger obligation. In this manner, present a strategy which can keep up and use this relationship to speed the inquiry stage is interesting. Distributed computing is the open source stage, expansive system access, on interest capacity, fast flexibility are the few favorable circumstances of distributed computing yet security and protection are the serious issue. In existing clarification information are put away in plain message because of which information is defenseless for interruption.

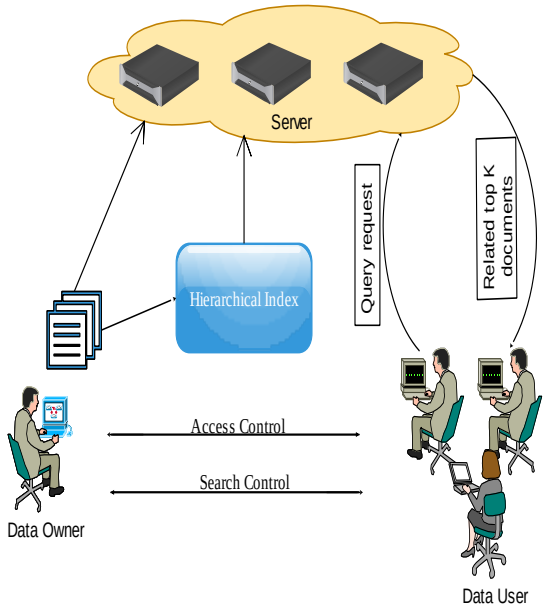

Fig - 1: Architecture of cipher text search

A vector space model is utilized and each archive is spoken to as a vector, which means each record can be viewed as a point in a high dimensional space.

2. LITERATURE REVIEW

The benefit of capacity as an administration numerous ventures are moving their profitable information to the cloud starting with one area then onto the next, since it has less cost, effectively versatile and can be gotten to from anyplace whenever [1]. Here security as a parameter is utilized to build up trust. Cryptography is one method for building up certitude. Cryptographic strategy is utilized for accessible encryption for giving security. Numerous specialists have been taking a shot at creating productive accessible encryption plans. This paper investigates some powerful cryptographic strategies dependent on information structures. CRSA and B-Tree is utilized to build the degree of security. It attempted to contraption the hunt on scrambled information utilizing cloud stage [1].

Distributed computing is to give answer for information redistribute and top notch information administrations. More establishments, associations and companies are dissecting the probability of having their applications, information and their IT resources in cloud [2]. As the information and there cloud's size builds, looking of the

applicable information is ordinary to be a danger. To defeat this risk, search file is made to help in quicker search. Notwithstanding, search Index creation and its estimation has been intricate and tedious, prompting cloud-down time there by thwarting the quickness in responding to information demand for mission basic prerequisites. Focal point of this paper is to clarify how reusability of pursuit list is diminishing the unpredictability of hunt list com put activity. Search record is proposed to be made utilizing parameters like closeness importance, client positioning and plan vigor. Client positioning ensures a keyword is utilized every now and again in the transferred information [2].

The proposed framework characterized that the reusability of inquiry list idea decreases cloud expending time while keeping up the security utilizing accessible symmetric encryption (SSE).The record mentioned from client is brought from the cloud, utilizing Two-round accessible encryption (TRSE) plot that supports top-k multi-keyword recovery [2].

These days, more individuals are roused to re-appropriate their nearby information to open cloud servers for extraordinary comfort and decreased expenses in information the board. Be that as it may, in light of securities issues, delicate information ought to be scrambled before re-appropriating, which resign conventional information use like catchphrase based record recovery. A protected and proficient multi keyword positioned search conspires over scrambled information, which furthermore underpins dynamic update tasks like cancellation and inclusion of records [3].

A list tree dependent on vector space model is utilized to give multi catchphrase search. Cosine closeness measure is utilized to help exact positioning for item. "Voracious first Traverse Strategy" search calculation is proposed to improve search effectiveness. Likewise utilize different security necessities in the known figure content risk model for giving insurance to the catchphrase. Trials on this present reality dataset demonstrate the viability and proficiency of proposed plot [3].

3. PROPOSED SYSTEM

The issue of keeping up the cozy connection between various plain reports over an encoded area has been search and proposes a bunching strategy to take care of this issue. A pursuit methodology is utilized to upgrade the rank protection. This inquiry procedure receives the backtracking calculation to comprehend any question where it happens. By applying the Merkle hash tree and cryptographic mark is utilized for tree structure confirmation and furthermore gives a check system to guarantee the rightness and culmination of query items.

4. PROBLEM DEFINITION

4.1 Threat Model

The adverse ability is concluded in two threat models. Known cipher text model: In this model, Cloud server can get encrypted document collection, encrypted query keywords and encrypted data index.

Known background model: In this model, cloud server knows more information than that in known cipher text model. Statistical background information of data set, such as the document frequency and term frequency information of a specific keyword, can be used by the cloud server to launch a statistical attack to infer or identify specific keyword in the query which further reveals the plain-text content of documents.

4.2 Design Goals

Search efficiency: The time complexity of search time of MRSE-HCI is less because the scheme needs to be logarithmic against the size of data collection in order to deal with the uncontrollable growth of document size in big data.

Retrieval accuracy: Retrieval precision is related to two factors: the relevance between the query and the documents in result set, and the relevance of documents in the result set. Integrity of the search result the correctness, completeness and freshness of the document should be maintained.

5. SYSTEM ARCHITECTURE AND ALGORITHM

The MRSE-HCI architecture is illustrated as below. The data owner is responsible for collecting documents, building document index and also redistribute them in an encrypted format to the cloud server. The cloud server provides a huge storage space, and the resources needed by cipher text search.

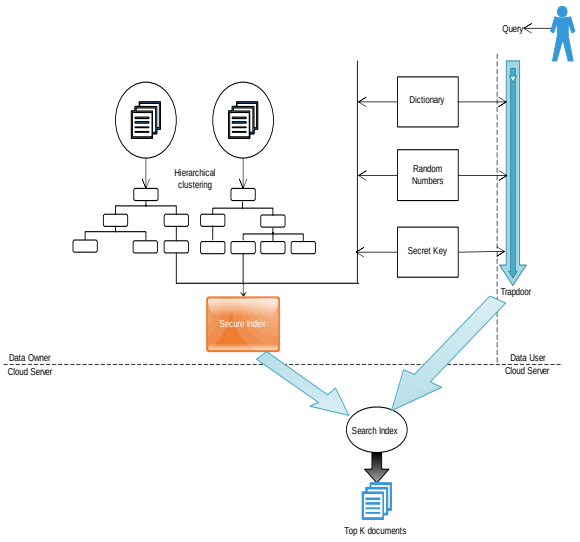

Fig – 2: MRSE-HCI architecture

After receiving a request from the data user, the cloud server searches the encrypted index, after getting the result then it sends back to top k documents. The data user properly chosen the number k and system aims to protecting data from leaking information to the cloud server for improving the efficiency of cipher text search. In this model, both the data owner and the data user are trusted, because the cloud server is semi-trusted, which is consistent with the architecture. The cloud server will strictly follow the predicted order of keywords and also try to get more information about the data and the document index. Before accessing the data, data owner needs to get authorization from the data owner.

Until now many hierarchical clustering methods have been proposed but all this method are not comparable to the partition clustering method , K-mean and K-Medoid are popular clustering algorithms but the size of k is fixed here. So this paper proposed a quality hierarchical clustering algorithm based on dynamic k-means.

Algorithm for Dynamic k-means

1. input the initial set of k cluster C
2. set the threshold TH_{min}
3. while k s not stable
4. Generate a new set of cluster center C_0 by k-means
5. For every cluster center C_0, i
6. Get the minimum relevance score: min (S_i)
7. If the min (S_i) <TH_{min}
8. Add new cluster center: k=k+1
9. go to while
10. Until k is steady.

Every cluster is checked whether its size exceeds TH or not. If the size exceeds the cluster will split into child cluster which is formed by dynamic k-means this procedure will be repeated until the entire cluster meet the requirement of maximum cluster size.

Algorithm Quality Hierarchical Clustering (QHC)

1. input document and set the size threshold TH.
2. build cluster ser C0 in first level by dynamic k-means.
3. while there new cluster set Ci.
4. for every cluster Ci,j.
5. if the size Ci,j is bigger than TH.
6. split this cluster into sub-cluster Ci+1.
7. until all clusters match the size constraint.

The retrieved document has possibility to be wrong because of unstable network and the data may damage due to hardware or software failure, so verifying the authenticity of search result is critical issue in cloud environment. Therefore, the minimum hash sub tree is designed to verify the correctness and freshness of search result.

Algorithm building minimum hash sub tree (MHST)

1. build hast tree dependent on various leveled grouping result.
2. for each leaf hub I.
3. calculate its hash esteem.
4. while not tree root.
5. for each non leaf hub j.
6. calculate its hash esteem.
7. construct hub (idj).
8. goto the upper level.
9. calculate tree root's hash esteem.
10. calculate the mark of hash esteem.
11. End

6. CONCLUSION

The maintaining of semantic relationship between different pain documents has been explored. This method is used to enhance the performance of the semantic search. MRSE-HCI architecture is used for the transformation to the requirements of data explosion, online information retrieval and the semantic search. A verifiable mechanism is proposed for the guarantee correctness and completeness of search results. To evaluate the search efficiency, accuracy and the rank privacy using experiments. Multi keyword based ranked search problem is not solved by using this proposed method. It also brings an improvement in search efficiency, rank security, and the relevance between retrieved documents.

REFERENCES

[1] Chi Chen, Peisong Shen, Jiankun Hu, Song Guo, Zahir Tari, Albert Y Zomaya, “An efficient privacy preserving ranked keyword search method” in Proc. IEEE Transactions on parallel and distributed systems vol-27, 2016.

[2] D. X. D. Song, D. Wagner and A. Perrig, "Practical techniques for searches on encrypted data", in Proc. IEEE Symp. Security Priv, BERKELEY, CA, PP: 44-55, 2000.
[3] D. Boneh, G. Di. Crescenzo, R. Ostrovsky and G. Persiano, "Public key encryption with keyword search", in Proc. EUROCRYPT, Interlaken, SWITZERLAND, pp: 506-522, 2004.