An Efficient Domain and Range Specific Multi-keyword Search Method

Miss.Sanchita R Ingawale[1], Dr. S. T. Singh[2]
[1]Reserach Scholar,Department of Computer Engineering,
P K Technical Campus,Pune,India.
[2]Professor,P K Technical Campus,Pune,India.

Abstract: A standout amongst the most central administrations of distributed computing is Cloud stockpiling administration. Immense measure of delicate information is put away in the cloud for simple remote get to and to diminish the cost of capacity. It is important to scramble the delicate information before transferring to the cloud server so as to keep up protection and security. The Domain and Range Specific Multi-watchword Search (DRSMS) plan was proposed to minimize the pursuit time and Index storage room. This plan embraces gathering sort strategy to part the record document into D Domains and R Ranges. The Domain depends on the length of the watchword; the Range parts inside the area in light of the primary letter of the catchphrase. A numerical model is utilized to look over the scrambled recorded watchword that takes out the data spillage.

1. Introduction

Cloud computing has turned into an imperative arrangement stage for appropriated applications particularly as information stockpiling and data administration benefit because of its huge potential in registering, stockpiling and different applications. From a joint pool of registering assets that are configurable, it permit stockpiling of remote information, on request utilization. A versatile and monetary arrangement is given by distributed computing framework to overseeing data and sharing assets. Framework upkeep overhead and equipment programming consumption is lessened by it. It offer suitable correspondence way to share assets between information proprietors and information clients. The prevalence of cloud administrations, for example, Microsoft Azure, AWS Amazon Services, Apple iCloud, Google AppEngine, has empowered organizations to move their information onto the cloud. The information proprietor can convey the individual data onto open cloud and information client can get to the data whenever and anyplace. Especially, tremendous measure of data and workloads can be sent by end-client to the cloud. Utilization of boundless figuring in a compensation for each utilization asset sharing model serve a one of the advantages and this allows the client to pay just for the measure of administration utilized.

The exceedingly challenge assignments confronted by Cloud Computing framework, information classification, unwavering quality and security concerns involve the principle position. In viable, general society cloud which are far from the trusted spaces contain the classified information. The information transferred by the information proprietors to the cloud bring worry of conceivable information misfortune, unscrupulous use of secret information as the proprietors don't have any immediate control over the delicate data. By and large, cloud servers are named as inquisitive and untrusted substances. Information proprietor frustrate to actualize cloud advancements when an instance of break of data to outsider or cloud supplier itself is conceivable. Thus giving abundant security and secrecy assurance to data that is helpless to rupture is of high significance. This gets utilized in application intended for social insurance, money related and government information. In order to keep the rupture of more private information that is transferred to the cloud, data is encoded in advance and after that transferred to the cloud server. To recover information records, conventional searchable symmetric encryption (SSE) method relies on upon watchword look instrument yet they bolster just Boolean catchphrase seek with no affirmation of n document recovery precision. This instrument is wasteful in recovery; it requests a lot of post-handling overhead and brings about superfluous system movement.

To settle the issues of information rupture in cloud, ebb and flow arrangements utilize the accompanying ways to deal with give looking capacity on cloud information on premise of watchwords. A gathering of catchphrases are distinguished and put away on the record document. For each document a list vector is outlined. After the production of record vectors, all the file vector are converged in a file document and delivered. The record document hence created and the information document are transferred to cloud servers after encryption. Presently, the data is set up to permit inquiries from the information client Cipher-content is upheld by cloud servers based n questions as takes after. A catchphrase construct seek inquiry in light of the cloud containing the scrambled information is sent by the information client and watchwords that are encoded are sent to the cloud. In the wake of accepting the question, the cloud server actualizes as pursuit on the encoded record and returns aftereffects of the rundown of relatable documents. The information client then settles on a

decision of the records that are fundamental and are recovered from the cloud server. With the assistance of the approved mystery key, the client decodes the required encoded records that were recovered from the cloud. Thusly, insurance of information from break and information classification is shielded. Amid the whole strategy, plaintext data or catchphrases are undetectable to the cloud serversDomain and Range Specific Query.

2. Domain Range Multi-Keyword Search

Domain and Range Specific Multi-catchphrase Search (DRSMS) plan is proposed to minimize the inquiry time and Index storage room.

This plan receives accumulation sort strategy to part the file record into D Domains and R Ranges. The Domain depends on the length of the catchphrase; the Range parts inside the area in light of the principal letter of the watchword.

It performs viable and productive ordering and seeking on scrambled information. The calculation lessens record storage room and positioned searchable encryption time for top - k multi watchword recovery on cloud touchy data. Ordering stage decreases file calculation time and record storage room.
.

3. System Architecture

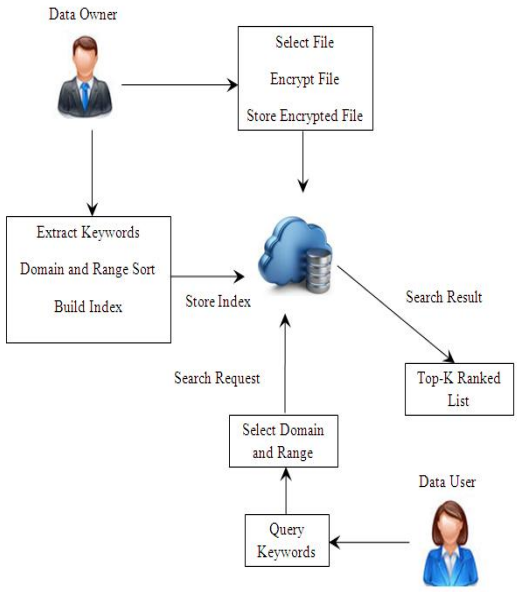

Fig1: System Architecture

This paper proposes a novel Index era and a questioned watchword seek system Domain and Range Specific Multi-catchphrase Search (DRSMS), that backings precise pursuit over encoded cloud information. DRSMS gives secure, productive and compelling indexed lists inside a brief span and it shields privately of information from the cloud specialist organization and unapproved clients.

Data Owner
The Data proprietor is an accumulation of n documents spoke to by F = (f1, f2, . . . , fn) to be outsourced on the cloud space. The terms are separated before outsourcing a record and a list document is fabricated. The list record contains: term, document ID(fi) and recurrence as area and range. It is anything but difficult to discover the watchword over scrambled list record. Encourage the file document and accumulations of n records are scrambled before outsourcing to the cloud.

Cloud Server
The cloud server speaks with the information proprietor and information client. It has capacity and recovery administrations for the outsider. The cloud supplier is not included in any erasure or change of information. In the vast majority of the SSE plots, the cloud server is considered as legitimate yet inquisitive, status to take in data from put away information.

Data User
The Secret key is utilized to produce the trapdoor tw between the Data User and Cloud Server. The approved client can send questioned catchphrase to the cloud server to look the top-k records. The questioned watchword look relies on upon area and scope of the file. Subsequent to accepting the inquiry catchphrase, the cloud server gives back the significant records as quick as could be expected under the circumstances identified with the questioned watchword. The information client can decrease the correspondence cost by sending the ideal esteem k. The top-k results are come back to the information client from the cloud server.

More details of System:
Setup(λ):
The parameter produces Secret Key (SK) and Public Key (PK) for the proposed plot. Information proprietor appropriates the mystery key to the approved clients.

Index_Generation(PK, F):
From the accumulation of delicate records F, each fi document separates the one of a kind watchword to build the searchable secure list I" by means of the encryption key (SK). Sorting depends on Domain D and Range R arrange; here Domain D is taken a length of the catchphrase and Range R is chosen subset inside the Domain D. The searchable record registers the catchphrases furthermore contains the recurrence and document IDs.

Trapdoor Generation(PK,REQ):
The questioned catchphrases ask for REQ creates secure trapdoor between the information client and the cloud server. The container trapdoor tw is worked from client's catchphrase ask for REQ and after that scrambled into a safe trapdoor tw with general society key (PK).

Search(I,Q):
The questioned catchphrase Q is figured by Equation 2 and contrasted and secure the searchable record I and returns the encoded frame beat k coordinating documents fi.

4. Proposed Algorithm Steps

This section describes structure and steps involved in implementation of algorithm used in the venture. These are listed and briefed as follows:

Algorithm-1 Build Index
Input: A Collection of n Data Files F = (f1, f2, . . . , fn)
Output: Domain and Range sorted Index file I''
for fi = 1 to F do
each file fi 2 F;
Scan F and Extract each term in fi, denoted as a
W = (w1,w2,w3, . . . ,wn,) ;
Normalized and remove the stop words from W;
for i = 1 to W do
count frequency of each word in fi;
store the hL||FLW||ID(fi)||wi||Si in I;
Index I sort based on length L of keyword and store in specific Domain Di;
Each Domain Di sort based on alphabets FLW and store in Specified Bucket Bij ;
for i = 1 to W do
Compute _(wi) for each keyword wi
Each computed results stores hID(fi)||_(wi)||Si in ascending order of index I';
I'' = encryption of Index file I';
return I;

Algorithm-2 Search Query
Input: Queried Keywords Q
Output: Top-k Search Result IDlist
Procedure: Search Query(K,Q)
Search keyword taken as hFLW||qi||Li for each word;
Compute _(qi) for each Queried keyword qi
Di is initialised to 3;
while L = Di do
Select the Domain Di depends on the length of the keyword out of Dn;
if L > Dmax then
length not found
else
Di + +
Procedure: BinarySearch(D[ ], FLW,Ds,De)
return range Rij select with in the Domain Di based on the first character of the search keyword out of Rnm;
for i= Cs to Ce do
if (_(qi) = _(wi)) then
Retrieve the file ID(fi);
else
No match found;
for i = 1 to ID(fn) do
retrieve the Score S for each file ID(fi);
Sort the Score S list in descending order;
Retrieve the top-k IDlist from the Index;
Decrypt the User interested file by using Top-k list;

5. Result and Performance Analysis

Our proposed system solves the problem of storage computational overhead for storing index of large set of documents D across cloud platform with N users. For performance measure we compare the computational overhead that is incorporated in implementing the optimized index storage for large clod database of documents D. Computational overhead is involved in process of index storage which is measured in terms of Storage cost and Search time cost required to generate index for document D uploaded by N users .

Figure 1 shows the execution time of existing and proposed methods- The proposed method is DRMS method of searching which is used avoid unwanted search and thus time required to execute is very less than the time required to execute existing system- We have computed time by subtracting start time from end time for Trapdoor with 3 Keywords-

As we can see as No of Keywords in trapdoor increases time to search the same also increases where as for proposed time is almost constant / increasing slowly-

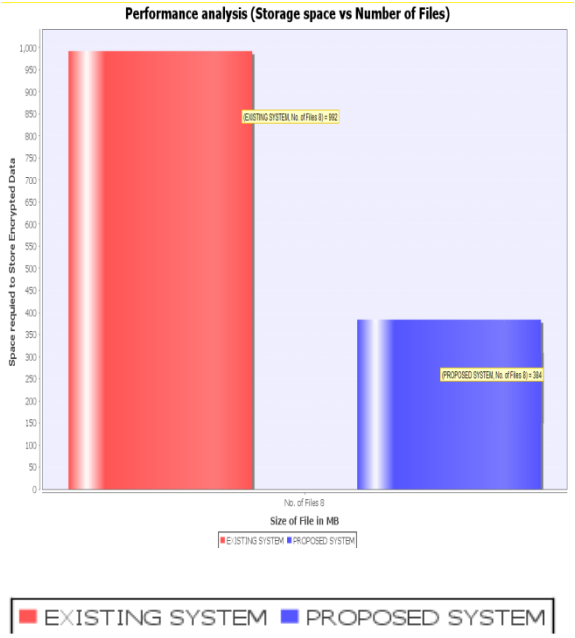

Figure 2: Storage Space vs- Number of Files

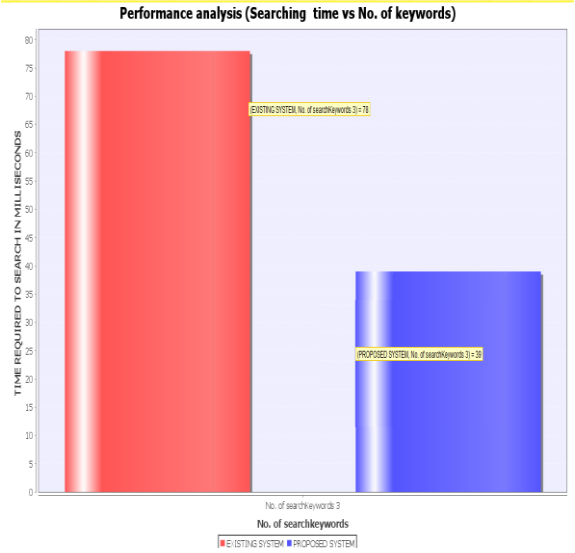

Figure 3: Search time vs- No- of search Keywords

Figure 3 shows the amount of storage of keywords vs- Number of Files as Number of files increases the amount of storage also increases as common keywords are also stored multiple times in index where as in proposed we store keywords on once divided in to Domain and range- Hence less amount of storage required to store keywords for N files- Here N= 8 files.

5. Conclusion

This paper presents another multi-keyword look procedure for accomplishing successful information usage through Internet over remotely put away scrambled cloud information- RSSE plan is wasteful to accomplish multi-catchphrase positioned seek on a huge dataset- To defeat the disadvantage in RSSE plot, we propose another Domain and Range Specific Multi catchphrase Search (DRSMS) calculation that backings more proficient and precise inquiry- DRSMS calculation performs powerful and productive ordering and looking on encoded information- The calculation diminishes file storage room and positioned searchable encryption time for top-k multi catchphrase recovery on cloud touchy data.

.

References:

[1] L. Chen, X. Sun, Z. Xia, and Q. Liu, "An Efficient and Privacy- Preserving Semantic Multi-Keyword Ranked Search over Encrypted Cloud Data," on International Journal of Security and Its Applications, vol. 8, no. 2, pp. 323–332, 2014.
[2] K. Li, W. Zhang, K. Tian, R. Liu, and N. Yu, "An Efficient Multikeyword Ranked Retrieval Scheme with Johnson-Lindenstrauss Transform over Encrypted CloudData," in Proceedings of the International Conference on Cloud Computing and Big Data (CloudCom-Asia), pp. 320–327, 2013.
[3] W. Zhang, S. Xiao, Y. Lin, T. Zhou, and S. Zhou, "Secure Ranked Multi-keyword Search for Multiple Data Owners in Cloud Computing," in Proceedings of the 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 276–286, 2014.
[4] M. Li, S. Yu, N. Cao, and L. Wenjing, "Authorized Private Keyword Search over Encrypted Data in Cloud Computing," in Proceedings of the 31st International Conference on Distributed Computing Systems (ICDCS), pp. 383–392, 2011.
[5] S. Buyrukbilen and S. Bakiras, "Privacy-Preserving Ranked Search on Public-Key Encrypted Data," in Proceedings of the 2013 IEEE International Conference on High Performance Computing and Communications, pp. 165–174, 2013.
[6] B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-Preserving Multikeyword Fuzzy Search over Encrypted Data in the Cloud," in Proceedings of the 2014 International Conference on INFOCOM, pp. 2112– 2120, 2014.
[7] Y. Lu, "Privacy-preserving Logarithmic-time Search on Encrypted Data in Cloud." in 19th Annual Network and Distributed System Security Symposium, (NDSS), 2012.
[8] C. ¨Orencik and E. Savas¸, "Efficient and Secure Ranked Multi-keyword Search on Encrypted Cloud

Data," in Proceedings of the 2012 Joint EDBT/ICDT Workshops, pp. 186–195, 2012
[9] W. Wang, P. Xu, H. Li, and L. T. Yang, "Secure Hybrid-Indexed Search for High Efficiency over Keyword Searchable Ciphertexts," Future Generation Computer Systems, 2014.
[10] D. Wang, S. Fu, and M. Xu, "A Privacy-Preserving Fuzzy Keyword Search Scheme over Encrypted Cloud Data," in Proceedings of the 5th International Conference on Cloud Computing Technology and Science (CloudCom), vol. 1, pp. 663–670, 2013.
[11] C. Gu, Y. Guang, Y. Zhu, and Y. Zheng, "Public Key Encryption with Keyword Search from Lattices," in International Journal of Information Technology, vol. 19, no. 1, pp. 1–10,
[12] O. Goldreich and R. Ostrovsky, "Software Protection and Simulation on Oblivious RAMs," in Journal of the ACM (JACM), vol. 43, no. 3, pp. 431–473, 1996.
[13] Z. Xia, Y. Zhu, X. Sun, and J. Wang, "A Similarity Search Scheme over Encrypted Cloud Images based on Secure Transformation," in International Journal of Future Generation Communication & Networking, vol. 6, no. 6, pp. 71–80, 2013.
[14] M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient Similarity Search over Encrypted Data," in Proceedings of the 28th International Conference on Data Engineering (ICDE), pp. 1156–1167, 2012.
.