Impact Factor (SJIF): 3.632

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

Volume 3, Issue 10, October 2016

Cluster based multi-keyword search on cloud data.

MR.AMARNATH POHANERKAR

MCA DEPARTMENT JNEC AURANGABAD (M.S-431001)

Abstract —Cloud computing involves delivery of services to the internet. These are Iass(Infrastructure as a service),Paas(Platform as aservice), Saas(Software as a service) respectively. Provision of memory resource is a significant part of cloud computing. Now days because of implementation of ERP system or automization of industrial work many organization encounters limited storage space. As a result Organization has started outsourcing of data. Consequently, confidentiality of data and faster retrieval of data erupted as a major challenge. To best of my knowledge existing system for retrieval of huge amount of data is inefficient and time consuming. We have proposed cluster based multi-keyword search on cloud data that culminates in comparisons required to perform search are reduced and time required to search is reduced.

Keywords-component;
Cloud Computing, cluster generation phase, index generation phase, retrieval phase, faster access to cloud data.

I. INTRODUCTION:

Computing means the availing of computer as hardware or software resource. Cloud computing is defined as delivery of computing resources over internet. Cloud is based on pay per use model where the users are charged only for duration when the services are used. Cloud provides infinite storage to the users at limited setup and usage cost. The prime requirement of cloud usage is the internet availability. Due to the low cost and fast speed availability of internet, organizations are motivated to outsource their data on the cloud. There are large number of cloud service providers namely VMware, Microsoft, Google, Saleforce.com, Rackspace and Amazon. The organization can deploy a private cloud or may use a public cloud to store their data. As public cloud involves less cost and time, many organizations prefer using a public cloud.

The use of public cloud introduces security breaches such as data leakage, data theft. In order to provide security, the data is encrypted before outsourcing it to the cloud. So, confidentiality of the data is retained using cryptography. The use of cryptography to convert this confidential data into human unreadable form introduces the challenge of effective searching over this data. A native approach to search data is to download the entire encrypted dataset from the remote cloud server to the local machine. The entire dataset is decrypted and then the desired documents are retrieved. End users possessing mobile devices or thin clients are limited by the memory available and thus makes this approach inefficient for perform searching on encrypted data. In this paper, the aim is to provide a cluster based search scheme using which the desired documents can be retrieved with fewer comparisons. In the cluster based search scheme, the entire document collection is partitioned into multiple clusters to provide efficient searching. In this way number of comparisons is reduced, consequently, the average search time is also reduced.

II. DESIGN GOALS

In this paper, we propose a cluster based approach for multi-keyword search on encrypted cloud data. The goals of the proposed searched scheme are

1. To retrieve the desired relevant documents corresponding to the search query in an efficient manner by reducing the number of comparisons and time required.
2. To declare a search unsuccessful in an efficient manner by performing fewer comparisons and in minimum possible time.

III. POPOSED CLUSTER-BASED MULTI-KEYWORD SEARCH ON CLOUD DATA

Proposed cluster based multi keyword search on cloud data can be classified into three phases, namely, cluster generation, indexing and retrieval phase.

3.1 Cluster generation Phase

Initially keywords are extracted from each document. Based on the similarity of keywords extracted from each document, the documents are partitioned into multiple clusters. For instance, if an organization is willing to outsource their confidential data on cloud, then such document can be clustered based on categories such as Finance, inventory and personnel.

3.2 Indexing Phase

3.2.1 Document index generation

The keywords extracted from each document in the previous phase are used to generate the document index. For each keyword wi appearing with in the document, the secret key for HMAC(hash-based message authentication code) is generated using hash function(for example MD-5, SHA-1, and SHA-2). The hash value calculated on the given keyword is send to the data owner. The data owner retrieves the secret key corresponding to the hash value. The secret key is shared with the end user using public key encryption scheme. For keywords generating same hash value, the secret key is retrieved only once from the data owner.

Upon receiving the secret key, HMAC is calculated on the keyword to generate the hexadecimal index. This hexadecimal value is converted to binary equivalent which is reduced in length. This process of reduction of the index involves dividing the binary string into smaller substrings of equal length. If all bits in the substrings are zero, then the output value for that substring is zero. If any one of the bit is 1, then the output is 1.
The reduction step is
Example: - H (keyword) = 100 output is 1
H (keyword) = 111 output is 1
H (keyword) = 010 output is 1
H (keyword) = 011 output is 1
H (keyword) = 000 output is 0
The final index is obtained by taking the bitwise product of the indices obtained for each keyword

Index (keyword 1) =111….. 10
Index (keyword 1) =101….. 11
Index (keyword 1) = 101…..10

Algorithm -1: Document Index Generation.
Input: F: the document collection
for each document Fi belongs to F do
for each keyword Fi belongs to Fi do
secret_index ← hash(wi)
retrieve the secret_key corresponding to the secret_index from the data owner
index ← HMAC(wi, secret_key)
Ii ← Reduce(index)
end for
Document Index I←Bitwise product (Ii)
end for
return document index I

3.2.2 Cluster index generation

During cluster generation phase, the documents are partitioned into multiple clusters. These documents are used to generate the cluster index. The cluster index is generated using the bitwise product of the indices of all documents appearing with in the cluster.

3.2.3 Document Encryption

Depending on the choice of data owner, any symmetric key encryption algorithm can be used. Symmetric key based approach is used to handle large size data and is fast. For security this key is kept confidential.

3.3 Retrieval phase

3.3.1 Query Generation

The authorized user willing to perform search on the encrypted cloud data calculates the hash value for each search term is obtained from the data owner. Using the received secret keys, HMAC is calculated and the process similar to the cluster index generation is used to generate search query.

Algorithm 2:-
Query index generation

Input:{k1,k2,….kn}: set of keywords.
for each keyword ki do
secret_index←hash(wi)
if(secret key corresponding to the secret_index not previously received)
retrieve the secret_key corresponding to the secret_index from the data owner
end if
index←HMAC(wi,secret_key)
Ii←Reduce(index)
end for
Query IndexQ←Bitwise Product(Ii)
return Query Index Q

3.3.2 Document searching on the cloud server:

Upon receiving the query string, the cloud sever will select the appropriate cluster by comparing the query string with the cluster index. The comparison is made by comparing the bit positions with 0 values in the query index with the corresponding bit position in the cluster index. If both values are zero, then matching process will continue else it is assumed to be mismatch.

Algorithm -3: Cluster Selection
Input: Query String Q

for each cluster index Ii do
if for all the bits j with Qj=0, the value of Ii is also 0
return cluster i
end if
end for

3.3.3 Document Decryption:

The metadata corresponding to the documents retrieved as relevant is presented to the user. The end user analyses the metadata and requests cloud server for a particular encrypted document. In order to decrypt the document, secret key is required by user. A request is made to data owner to provide secret key. Upon receiving the secret key, the document is decrypted using the same algorithm which is used by data owner during document encryption.

IV. CONCLUSION

In this paper, we have proposed a cluster based approach for multi-keyword search on encrypted cloud data. The proposed scheme permits the user to efficiently perform search over the encrypted cloud data. To do so, thee data owner generates the cluster index and document index. The documents are encrypted and outsourced to cloud. Following things are analyzed:-

1. The proposed search scheme reduces the time and number of comparisons required to retrieve the desired documents.
2. The proposed search scheme requires less time as compared to the existing method even if documents appear in multiple clusters.

REFERENCES

[1] "CLOUD COMPUTING", HTTP:// EN.WIKIPEDIA.ORG/WIKI/CLOUD_COMPUTING.
[2] "The NIST definition of cloud computing", http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf
[3] "Top-10 cloud service providers",http://searchcloudcomputing.techtarget.com/photostory/2240149038/top-cloud-providers-of-2012/1/introduction.
[4] Morgan et al., "Factors affecting the adoption of cloud computing an exploratory study", http://www.staff.science.uu.nl/~vlaan107/ecis/files/ECIS2013-0710-paper.pdf.
[5]Ning Cao et al., "Privacy preserving multi-keyword ranked search over encrypted cloud data", in IEEE transactions on parallel and distributed systems,pp.222-233,2014.
[6]D song et al., "Practical techniques for searches on encrypted data", in Proc. of IEEE Symp. On Security and Privacy'00,Berkeley, CA, pp. 44-55,2000.
[7]Mehmet Ucal, "Searching on encrypted data", http://www.researchgate.net/publication/228757457_Searching_on_Encrypted_Data.
[8]E Goh, "Secure indexes", in Cryptology ePrint Archive, Report, 2003/216, http://eprint./iacr.org/