International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 03 | June-2015 www.irjet.net p-ISSN: 2395-0072

A study on keyword searchable frameworks for efficient data utilization in cloud storage

Bhagyashree M V[1], Pushpalatha M N[2]

[1] *Student, Dept. of ISE, M S Ramaiah institute of technology, Karnataka, India*

[2]*Asst. professor, Dept. of ISE, M S Ramaiah institute of technology, Karnataka, India*

Abstract: *Yester year advances in computer technology has resulted in huge success of cloud computing by gaining the software, platform and infrastructure from the vendor or third party. When it comes to data storage, cloud storage becomes the first choice. Public data centres stores user data and the hopeful benefit of cloud computing is outsourcing of data. Cloud computing paradigm facilitates data management for users by economically saving their per-capita investment. However there is a substantial problem of outsourcing the data for accessing unauthorized data and hence there is no sense it is not effectively utilized. The biggest obstacle is how to attain efficient data utilization from public cloud storage focusing at various searchable techniques for improved data utilization. An effort is made in this paper to study different searching techniques for efficient data utilization from public cloud storage and further discussed in detail.*

Keywords: cloud server(CS), efficient data utilization, keyword search, data outsourcing, security

1. INTRODUCTION

As a simple and straight forward definition for cloud computing " an alternative of using local servers or personal computers to store, process or manage data is using a collection of remote servers hosted on internet"[1,2]. Outside cloud customers are provided with on demand, dynamically scalable computing power, services, storage and platforms through the internet. Third party service providers or vendors run the public cloud. Applications from number of users are combined together on servers, storages and networks. Cloud storage benefits users with on demand storage space and make user convenient, fool-proof and timely data acquisition.

The structuring of the rest of the paper is as follows. In sector 2, the existing searchable techniques are discussed. Sector 3, consists of the constraints and drawbacks. The performance analysis of all searching strategies is discussed in Sector 4. And finally, the conclusion is followed in sector 5.

2. KEYWORD SEARCHING TECHNIQUES

2.1 Authorized Private keyword Search (APKS)

Ming Li [4,5] proposed APKS and APKS+. APKS method upgrades the search efficiency using attribute hierarchy, and APKS+ magnifies the query privacy with the aid of proxy servers that swamp the dictionary storm (attacks). To the best of the knowledge APKS+ is the first most to achieve multi-dim range query and capacity delegation. The three important highlights of APKS is that provides keyword, Index and Query Privacy, Fine-grained Search Authorization, Multi-dimensional Keyword Search, Scalability and Efficiency.

2.2 Secured and privacy preserved keyword search

The cloud service providers usually allowed to participate in partial decipherment in order to reduce overhead caused due to computations. This framework was proposed by Qin Liu[3]. It provides both keyword and data privacy by public key encryption. The encrypted keyword trapdoor is submitted by the data user by using user private key to CS securely and receives the documents in encrypted form. And then decrypts it. This framework enables cloud provider to compute whether an electronic mail incorporated with the keywords as per the user.

1.1 Secured fuzzy keyword Search

This framework returns the matching files when data user search inputs accurately matches the already defined keywords based on the semantics, when there is a fail in exact match. Kui Ren[6] proposed this approach with symmetric searchable encryption (SSE). For keyword w_e, it builds the fuzzy keyword set Tw with the parameter called edit distance 'd' without affecting search correctness.

When user searches for a w_e with cloud server, it searches Tw and replies back with encrypted docs identical to Tw. They build the storage capable fuzzy keyword set by using wild card and fuzzy searchable index.

1.2 Secure and Efficient Ranked Keyword Search

Proposed by Cong Wang[7]. This framework improves the efficiency of searching. It retrieves the documents based on conceptual entities. It also uses a mechanism called topic detection and tracking. This technique solves the problem caused due to processing overhead, data and keyword privacy. The data owner constructs index along with the keyword's frequency-based relevance scores for documents. User request 'w_e' to CS with optional 'k' as Tw using the private key. The CS searches the document index with scores and replies with the encrypted documents depending upon the ranked order. Here order preserving symmetric encryption is made use of in order to safeguard cloud data. It efficiently uses outsourced files by providing inter cloud communication between data users and owners.

1.3 K-gram based fuzzy keyword Ranked Search

Previous cloud computing searchable encryption techniques allows users to search encrypted data by keywords securely, but these methods only support exact keyword search and will fail to perform if there are any spelling mistakes/morphological variants of words. Wei Zhou[8] proposed this method. A concept of k-grams index is used. K-grams is a sequence of k characters. For example, "dia", "ram", "agr" and "iag" are all the 3-grams of the word "diagram". Here owner creates k-gram fuzzy keyword index I_n for D number of documents and tuple as $<I_n,D>$ and is uploaded to CS. The encrypted doc D is uploaded to storage server. When keyword K is submitted by the data user, the k-gram fuzzy keyword set and computes weight of each word in the set and are searched with index I_n by CS. Then the CS displays all documents in sorted sequence identical to the index based on rank. Finally performance complexity is O(N) Where, N is total number of keywords.

1.4 Privacy-preserving Multi-keyword Text Search

Wen hai[9] Sun suggested this method which implements similarity based search resulting ranking, keyword privacy, Index and Query confidentiality and Query Unlinkability. Vector space model concept is employed here in order to build the index for encrypted files. It supports both conjunctive and disjunctive file search. The searchable index is created using Multidimensional search tree. Data owner creates encrypted query vector $\bar{Q}$ for file's keyword set. Finally CS searches index by MD algorithm and compares cosine measure of document and query vector and consequently returns top k encrypted files to user.

1.5 VFKS-Verifiable fuzzy keyword search

It maintains verifiability of search results in addition with fuzzy keyword search. Jianfeng Wang proposed this search strategy. He made use of symbol-tree and index $\bar{G}w$ which is having the unique value "proof" and the path for each node without key 'k'. When CS gets the keyword as the query, it searches $\bar{G}w$ and returns already stored encrypted documents. Finally the confirmation is done by users by cross-verifying the proofset and IDset generated from index. For each of the user query, the verifying cost (computation cost)is only a constant complexity.

2.8 Public-Key Encryption with Keyword Search (PEKS)

PEKS is semantically secure. This strategy was suggested by D. Boneh[10], which incorporates that CS contains encrypted files and keyword. User creates keyword trapdoor Tw using its private key to search W. The CS examines Tw with current existing encrypted keyword and sends encrypted file that is identical to it. There exists a secured channel between the data owner, cloud server and data user because owner does file encryption and server does the user authentication.

2.9 Secured multikeyword (ranked)Top-k Retrieval Search

Jiadi[11] suggests this search technique using Two-round searchable encryption (TRSE). Basically in first round, users submits multiple keyword requests REQ W' as encrypted query for acquiring data, keyword privacy and builds trapdoor(REQ, PK) as Tw and gives to CS. Then CS computes the scores from encrypted index for docs and gives the encrypted search result vector to the data user. And in the second round, user carries out decryption of N with secret key and calculates the doc ranking and then requests files with top-k(relavant) scores. The scoring and ranking of files is done on server side and user side respectively.

2.10 Attribute-based Keyword Search

It is a novel cryptographic solution. It enforces access control policies via means of cryptography. It decrypts cipher text that was encrypted according to access control policy by entities with proper credentials. There are two variants: key-policy ABE (KPABE) where the decryption key is associated to the access control policy, and

ciphertext-policy ABE (C-ABE) where the ciphertext is associated to the access control policy. And important feature is a cheating cloud can be held accountable. According to this method a data owner can control the search of his/her outsourced encrypted data through an access control policy and the authorized data users can outsource the search operations to the cloud server and force the cloud to trustfully execute the search operations.

3. CONSTRAINTS AND DRAWBACKS OF THE ABOVE DISCUSSED KEYWORD SEARCHING FRAMEWORKS.

3.1 APKS-Authorized Private keyword Search: In tradition, not all the attributes are hierarchical.

3.2 Secured and privacy preserved keyword search: The communication and computational cost for both decryption and encryption is heavier and amassed.

3.3 Secured fuzzy keyword Search: Doesn't support fuzzy search with public key-based searchable encryption, does not perform multiple keywords and semantic search, the updates for fuzzy searchable index are in-efficiently.

3.4 Secure and Efficient Ranked Keyword Search: multiple keyword searches not performed, little amount of overhead in index creation.

3.5 Verifiable fuzzy keyword search: this method, Verifiable fuzzy keyword search requires extra storage for storing the symbol tree fuzzy searchable index $\bar{G}w$, The updates for fuzzy searchable index is not so efficient.

3.6 Privacy assured searchable cloud Storage: This scheme exposes the approach pattern to the cloud server, It does not shields the sum of multiple keywords scores from the cloud server, which results in accessing the statistical data for re-identifying the search keywords, public key based searchable encryption is not supported.

3.7 K-gram based fuzzy keyword Ranked Search: The length of the k-gram based fuzzy keyword set purely depends on the jaccard coefficient value.

3.8 Privacy-preserving Multi-keyword Text Search: The similarity rank score of the document vector purely confide on the category of the document

3.9 Secure Multi-keyword Top-k Retrieval Search: Though the reduction and compression is utilized to decrease cipher text size, the key length is still too huge, The communication overhead will be too high if the encrypted trapdoor size is too large, It does not make effective searchable index update.

3.10 Public-Key Encryption with Keyword Search: building a secure channel is more expensive and inefficient, the trapdoor must be built for each keyword by the data user, it does not support multiple keyword search, Keywords may be hacked by KGA-Keyword Guessing Attack.

4. APPENDIX A – PERFORMANCE ANALYSIS OF SEARCHING METHODS

Table 1: performance analysis

Sl.no.	Searching methods	Performance complexity
1	Secure and privacy Preserving keyword search	Computation cost of O(Time(A))
2	Authorized Private Keyword Search (APKS)	While, N is total number of keywords and M is maximum size of the keyword set Setup= O(N2) Encryption=O(N) Search=O(M log N)
3	Secure and Efficient Ranked Keyword Search	O(log M) Where, M is domain score of keyword W
4	Secured fuzzy keyword search	While W is keyword, N the total number of keywords and M the maximum size of the fuzzy keyword Fuzzy set cost - O(|W|) Storage cost - O(MN) Search cost O(1)

5	Privacy assured searchable cloud Storage	O(|W|) Where W is the keyword
6	K-gram based fuzzy keyword Ranked Search	O(N) Where, N is total number of keywords.
7	Verifiable fuzzy keyword search (VFKS)	Storage cost - O(MN) Search cost O(1) Verify cost - O(1) Where, N is the total number of keywords M is the maximum size of the fuzzy Keyword
8	Privacy preserving Multi-keyword Text Search	O(|W|) Where W is the number of keyword
9	Public-Key Encryption with Keyword Search	Time cost of proxy =O(N) Communication cost = O(2T) Where, T is the average number of keywords searchable cipher text matching the query
10	Secure Multi keyword Top-k Retrieval Search	Setup= O(λ) Trapdoor=O(l) Score= O(Nl) Decryption =O(N)

5. CONCLUSION

This paper epitomizes distinct searching frameworks in the encrypted cloud data. We have pin-pointed and diagnosed the main issues that are to be satisfied for secured data utilization are keyword privacy, Data privacy, Index privacy, Query Privacy, Fine-grained Search, Scalability, Efficiency, Result ranking, Index confidentiality, Query confidentiality, Query unlinkability, semantic security and Trapdoor unlinkability. We have done a precise study on the security and data utilization issues in the cloud storage for some of the available searching strategies some of the searching techniques mainly targets on security and some on data utilization. The constraints of all the searching techniques are reviewed as well. Finally, by the atop survey, security can be provided by Public-Key Encryption and effective data utilization by fuzzy keyword search. We believe that this paper will help the researchers to shape their issues in the field of data utilization in cloud storage.

ACKNOWLEDGMENT

I would like to thank Dr. Vijaya Kumar B P, Head of Department of Information Science Engineering, MSRIT his valuable guidance.

REFERENCES

[1] J. Geelan. "Twenty one experts define cloud computing," Virtualization,August 2008.

[2] Foster et al.,"Cloud computing and grid computing 360-degree compared," Grid Computing Environments Workshop, 2008.GCE'08, 2009

[3] Qin Liuy, Guojun Wangyz, and Jie Wuz,"Secure and privacy preserving keyword searching for cloud storage services",ELSEVIER Journal of Network and computer Applications, March 2011

[4] Ming Li et al., "Toward Privacy-Assured and Searchable Cloud Data Storage Services", IEEE Transactions on Network, volume 27, Issue 4, July/August 2013

[5] Ming Li et al.," Authorized Private Keyword Search over Encrypted Data in Cloud Computing, IEEE proc. International conference on distributed computing systems,June 2011,pages 383-392

[6] Kui Ren et al.,, "Towards Secure And Effective Data utilization in Public Cloud", IEEE Transactions on Network, volume 26, Issue 6, November / December 2012

[7] Cong Wang et al.,"Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data", IEEE Transactions on parallel and distributed systems, vol. 23, no.8, August 2012

[8] Wei Zhou et al., "K-Gram Based Fuzzy Keyword Search over Encrypted Cloud Computing "Journal of Software Engineering and Applications, Scientific Research , Issue 6, Volume 29 32,January2013

[9] Wenhai Sun et al.,"Verifiable Privacy- Preserving Multi-keyword Text Search in the Cloud Supporting Similarity-based Ranking'', Accepted for IEEE Transactions on Parallel and Distributed Systems (TPDS)

[10] D. Boneh et al.,, "Public key encryption with keyword search," in Advances in Cryptology – EUROCRYPT 2004, Lecture Notes in Computer Science, vol. 3027, pp. 506-522, Interlaken, Switzerland, 2004. Springer

[11] Jiadi Yu, Peng Lu, Yanmin Zhu, Guangtao Xue, Member, IEEE Computer Society, and Minglu Li,"Toward Secure Multikeyword Topk Retrieval over Encrypted Cloud Data", IEEE Transactions on dependable and secure computing, vol. 10, no. 4, July/August 2013

BIOGRAPHIES

Bhagyashree M V received B.E (ISEI) in 2007 from Vidyavardhaka college of engg, mysore affiliated to VTU Belgaum. She is currently pursuing M.Tech in MSRIT affiliated to Vishveshwariah technological university, Belgaum. Her research interests include cloud computing, data mining and big data.

Mrs. Pushpalatha M N, born on 21st Jan 1983 is an assistant professor in Department of Information Science and Engineering at M S Ramaiah Institute of Technology, Bangalore-54. Her areas of interest is Software Engineering and data mining. She completed M.Tech in Computer Science and Engineering from M S Ramaiah Institute of Technology, Visvesvaraya Technological University.

