ISSN 2395-1621

Review Paper on Multi-keyword top-k search over encrypted cloud data

[#1]Tushar Mahajan, [#2]Shridhar Rodge, [#3]Sumeet Pawar, [#4]Ashish Mangukiya, [#5]Prof. D.A. Chaudhari

[1]tush974@gmail.com
[2]rodgeshridhar13@gmail.com
[3]sumeetpawar16000@gmail.com
[4]ashishmangukiya97@gmail.com
[5]dipalee.rane@gmail.com

[#12345]Department of Computer Engineering,
D.Y.Patil College of Engineering, Akurdi, Pune.

ABSTRACT

Now a days cloud computing has become more popular, so more information possessors store their data to cloud servers for great convenience and less monetary value in data management. Cloud storage provides a convenient, massive, and scalable storage at low cost, but data privacy is a major concern that prevents users from storing files on the cloud trustingly. One way of enhancing privacy from data owner point of view is to encrypt the files before outsourcing them onto the cloud and decrypt the files after downloading them. However, data encryption is a heavy overhead for the mobile devices, and data retrieval process becomes complicated communication between the data user and cloud. Important Information should be encrypted before outsourcing our personal data in public or third party. In proposed system the problem of a secure multi-keyword search on cloud is solved by using encryption of data before it actually used. Here, we used keyword based algorithm for searching the file search. Searching is done efficiently by using the keywords user can search his document by multiple keywords. The problem of a secure multi-keyword search on cloud data is solved by using encryption of data before it actually used or uploaded on cloud. The top-k multi-keyword search makes the process more efficient and faster than searching over single keyword search.

ARTICLE INFO

Article History

Received: 20th May 2018

Received in revised form :

20th May 2018

Accepted: 23rd May 2018

Published online :

23rd May 2018

Keywords: Multi-keyword search, Encryption, Ranking, Cloud, Key-exchange

I. INTRODUCTION

Cloud Computing is a new but increasingly mature model of enterprise IT infrastructure that provides storage and on-demand high quality applications and services from a shared pool of configuration computing resources. The cloud customers, individuals or enterprises, can outsource their local complex data system into the cloud to avoid the costs of building and maintaining a private storage infrastructure. However, some problems may be caused in this circumstance since the Cloud Service Provider (CSP) possesses full control of the outsourced data. Unauthorized operation on the outsourced data may exist on account of curiosity or profit. To protect the privacy of sensitive information, sensitive data (e.g., emails, photo albums, personal health records, financial records, etc.) should be encrypted by the data owner before outsourcing, which makes the traditional and efficient plaintext keyword search technique useless. The simple and awkward method of downloading all the data and decrypting locally is obviously impractical. So, two aspects should be concentrated on to explore privacy preserving effective search service. Firstly, ranked search, which can enable data users to find the most relevant information quickly, is a very important issue. Cloud computing is one way of computing. Here the computing resources are shared by many users. The benefits of cloud can be extended from individual users to organizations. The data storage in cloud is one among them. The virtualization of hardware and software resources in cloud nullifies the financial investment for owning the data warehouse and its maintenance. Many cloud platforms like Google Drive, iCloud, SkyDrive, Amazon S3, Dropbox and Microsoft Azure provide storage services. Security and privacy concerns have been the major challenges in cloud

computing. The hardware and software security mechanisms like firewalls etc. have been used by cloud provider. These solutions are not sufficient to protect data in cloud from unauthorized users because of low degree of transparency. Since the cloud user and the cloud provider are in the different trusted domain, the outsourced data may be exposed to the vulnerabilities. Thus, before storing the valuable data in cloud, the data needs to be encrypted. Data encryption assures the data confidentiality and integrity. To preserve the data privacy we need to design a searchable algorithm that works on encrypted data.

II. LITERATURE SURVEY

[2] The paper deals with the problem of privacy-preserving top-k keyword similarity search over outsourced cloud data. Taking edit distance as a measure of similarity, we first build up the similarity keyword sets for all the keywords in the data collection, then calculate the relevance scores of the elements in the similarity keyword sets by the widely used tf-idf theory.

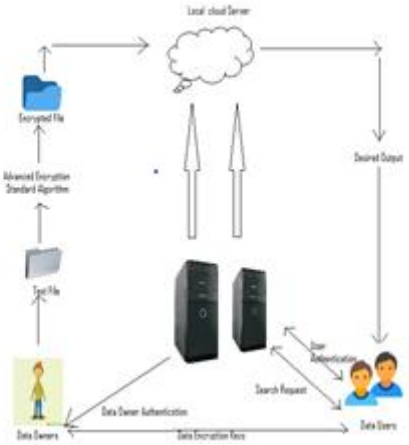
[3] This paper first propose a basic idea for the MRSE based on secure inner product computation, and then give two signifi- cantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models.

[4] We consider the issue of building a safe cloud storage services on top of an open cloud foundation where theservice provider is not totally trusted by the user. We depict, at an abnormal state, a few architectures that consolidate late and non-standard cryptographic primitives with a specific end goal to accomplish our objective. We review the benefits such a construction modeling would give to both customers and service providers and give an outline of late advances in cryptography roused specifically by cloud storage. We propose the first completely homomorphic encryption scheme, taking care of a focal open issue in cryptography. Such a plan permits one to figure subjective capacities over encypted data without the decoding key – i.e., given encryptions E(m1),...,E(mt) of m1,...,mt, one can efficiently process a smaller ciphertext that encrypts f(m1,...,mt) for any efficiently calculable capacity f. This issue was postured by Rivest et al. in 1978.

[5] Cloud storage provides a convenient, mas- sive, and scalable storage at low cost, but data privacy is a major concern that prevents users from storing files on the cloud trustingly.

[6] Specially investigate graph encryption method for an important graph query type, called top-k Nearest Keyword (kNK) searches. This work design several indexes to store necessary information for answering queries and guarantee that private information about the graph such as vertex identifiers, keywords and edges are encrypted or excluded.

III. PROPOSED SYSTEM



1. Keyword Expansion
To enhance the accuracy of search results, the keywords are removed from outsourced content documents required to be stretched out by regular synonyms or comparable words, as cloud customers, searching information may be the synonyms of the predefined keywords.

2. Upload Encrypted Data
After expansion of keywords the data owner assist data with encrypting the document utilizing AES Algorithm and after that upload the encrypted document to the cloud for storage reason. This permits data owner to store their secret key in extremely secure way without presenting it to the clients of framework. For this, secret key is put away again in encrypted frame.

3. Search Module
This module helps clients to enter their query keyword to get the most important documents from set of uploaded documents. This module recovers the documents from cloud which coordinates the query keyword.

4. Download Ranked Results
Clients can download the resultant arrangement of documents just if he/she is approved client who has allowed consent from data owner to download specific document. Owner will send encrypted secret key and session key to client to decrypt the document.

IV. AES AND DES COMPARISON

Data Encryption Standard becomes known as a common standard used for encryption of data around the world and forms a secret key cryptography that only has one key for the use of decryption. Advanced Encryption Standard becomes known as an exceptional standard employed around the world by the government of the United States to safeguard its secrets.

DES is a usage of a Feistel Cipher. It utilizes 16 round Feistel structure. The square size is 64-bit. However, key length is 64-bit, DES has a successful key length of 56 bits since 8 of the 64 bits of the key not utilized by the encryption calculation. The production of an NSA-endorsed encryption standard at the same time brought about its

speedy universal selection and across the board scholastic investigation. Contentions emerged out of characterized plan components, a short key length of the symmetric-key piece figure outline, and the contribution of the NSA, supporting doubts about an indirect access. The critical scholarly investigation the calculation gotten after some time prompted the cutting edge comprehension of square figures and their cryptanalysis. DES works by utilizing a similar key to scramble and decode a message, so both the sender and the beneficiary must know and use a same private key. Once the go-to, symmetric-key calculation for the encryption of electronic information, DES has been superseded by the more secure Advanced Encryption Standard (AES) estimate. Amid the most recent couple of years, cryptanalysis had discovered a few shortcomings in DES when core chose are feeble keys. These keys might get kept away. DES has turned out to be an extremely outlined square figure. There have been no huge cryptanalytic assaults on DES other than through fundamental inquiry.

The Advanced Encryption Standard, or AES, is a symmetric piece figure picked by the U.S. government to ensure grouped data and is actualized in programming and equipment all through the world to scramble touchy information. AES has been received by the U.S. government and presently utilized around the world. It supersedes the Data Encryption Standard (DES), which was distributed in 1977. The calculation portrayed by AES is a symmetric-key calculation, which means a similar key is utilized for both scrambling and unscrambling the information. This new propelled encryption calculation would be unclassified and must be "equipped for securing delicate government data well into the following century," as indicated by the NIST declaration of the procedure for improvement of a propelled encryption standard calculation. It was planned to be anything but difficult to actualize in equipment and programming, and in limited situations (for instance, in a shrewd card) and offer excellent guards against different assault methods. AES depends on an outline standard known as a substitution-stage arrange, a blend of both substitution and change, and is quick in both programming and equipment. Not at all like its antecedent DES, AES does not utilize a Feistel arrange. AES is a variation of Rijndael which has a settled piece size of 128 bits, and a key size of 128, 192, or 256 bits.

V. ALGORITHM

AES:

The Advanced Encryption Standard (AES) was published by the National Institute of Standards and Technology (NIST) in 2001. It is symmetric block cipher

AES is a block cipher intended to replace DES for commercial applications. It uses a 128-bit block size and a key size of 128, 192, or 256 bits. AES does not use a Feistel structure. Instead, each full round consists of four separate functions: byte substitution, permutation, arithmetic operations over a finite field, and XOR with a key.

The cipher takes a plaintext block size of 128 bits, or 16 bytes.The key length can be 16, 24, or 32 bytes (128, 192, or 256 bits). The algorithm is referred to as AES-128, AES-192, or AES-256, depending on the key length.The input to the encryption and decryption algorithms is a single 128-bit block.

The cipher consists of N rounds, where the number of rounds depends on the key length: 10 rounds for a 16-byte key, 12 rounds for a 24-byte key, and 14 rounds for a 32-byte key.

AES processes the entire data block as a single matrix during each round using substitutions and permutation.

Four different stages are used, one of permutation and three of substitution:

- **Substitute bytes:** Uses an S-box to perform a byte-by-byte substitution of the block
- **ShiftRows:** A simple permutation
- **MixColumns:** A substitution that makes use of arithmetic
- **AddRoundKey:** A simple bitwise XOR of the current block with a portion of the expanded key.

RSA:

Asymmetric encryption is a form of cryptosystem in which encryption and decryption are performed using the different keys—one a public key and one a private key. It is also known as public-key encryption.

The most widely used public-key cryptosystem is RSA. The difficulty of attacking RSA is based on the difficulty of finding the prime factors of a composite number.

RSA was developed in 1977 by Ron Rivest, Adi Shamir, and Len Adleman at MIT

The RSA scheme is a block cipher in which the plaintext and ciphertext are integers between 0 and *n* - 1 for some *n*.A typical size for *n* is 1024 bits, or 309 decimal digits. That is, *n* is less than 2^1024.

RSA Algorithm Flow:

Key Generation Alice

Select p, q	p and q both prime, $p \neq q$
Calculate $n = p \times q$	
Calcuate $\phi(n) = (p-1)(q-1)$	
Select integer e	$\gcd(\phi(n), e) = 1; 1 < e < \phi(n)$
Calculate d	$d \equiv e^{-1} \pmod{\phi(n)}$
Public key	$PU = \{e, n\}$
Private key	$PR = \{d, n\}$

Encryption by Bob with Alice's Public Key

Plaintext:	$M < n$
Ciphertext:	$C = M^e \bmod n$

Decryption by Alice with Alice's Public Key

Ciphertext:	C
Plaintext:	$M = C^d \bmod n$

VI. CONCLUSION

The proposed system concludes that, searching can be performed on the encrypted data without decrypting the whole data using above mentioned algorithms. The privacy of the user is maintained in an efficient manner.

REFRENCES

[1] X. Ding, P. Liu and H. Jin, "Privacy-Preserving Multi-keyword Top-k Similarity Search Over Encrypted Data," in IEEE Transactions on Dependable and Secure Computing. doi: 10.1109/TDSC.2017.2693969.

[2]TENG Yiping, CHENG Xiang, SU Sen, WANG Yulong, SHUANG Kai "PrivacyPreserving Top-k Keyword Similarity Search over Outsourced Cloud Data" in China Communications, vol. 12, no. 12, pp. 109-121, December 2015. doi: 10.1109/CC.2015.7385519

[3] Ning Cao, Cong Wang, Ming Li, Kui Ren, Wenjing Lou "Privacy-Preserving Multi- Keyword Ranked Search over Encrypted Cloud Data" in IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222-233, Jan. 2014. doi: 10.1109/TPDS.2013.45

[4] K. Ren, C. Wang, Q. Wang et al., "Security challenges for the public cloud," IEEE Internet Computing, vol. 16, no. 1, pp. 69–73, 2012.

[5] Jian Li, Member IEEE/ACM, Ruhui Ma, Haibing Guan "TEES: An Efficient Search Scheme over Encrypted Data on Mobile Cloud" in IEEE Transactions on Cloud Computing, vol. 5, no. 1, pp. 126-139, Jan.-March 1 2017. doi: 10.1109/TCC.2015.2398426

[6] S. Kamara and K. Lauter, "Cryptographic cloud storage," in Financial Cryptography and Data Security. Springer, 2010, pp. 136– 149.

[7] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford University, 2009.

[8] O. Goldreich and R. Ostrovsky, "Software protection and simulation on oblivious rams," Journal of the ACM (JACM), vol. 43, no. 3, pp. 431–473, 1996.

[9] D. Boneh, G. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advance in Cryptology Eurocrypt 2004.Springer,2004 pp. 506–522.

[10] K. Ren, and W. Lou "Verifiable Privacy-Preserving Multi-Keyword Text Search in the Cloud Supporting Similarity-Based Ranking"-2013.

[11] W. Zhang, S. Xiao, Y. Lin, T. Zhou, and S. Zhou, "Secure ranked multi-keyword search for multiple data owners in cloud computing," in Dependable Systems and Networks (DSN), 2014 44th Annual.

[12] IEEE/IFIP International Conference on. IEEE, 2014, pp. 276–286.B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud," in IEEE INFOCOM, 2014.

