IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

# An Efficient Search Technique for Document Retrieval from Encrypted Cloud

**R. Rupa Devi[1], R. Sunil Kumar[2]**

PG Scholar, Department of CSE, MVGR College of Engineering, Vizianagaram, India[1]

Assistant Professor, Department of CSE, MVGR College of Engineering, Vizianagaram, India[2]

**Abstract:** Cloud computing, individuals can store their data on remote servers and allow data access to public users through the cloud servers. As the outsourced data are likely to contain sensitive privacy information, they are typically encrypted before uploaded to the cloud. This, however, significantly limits the usability of outsourced data due to the difficult of searching over the encrypted data. An efficient search technique for document retrieval from encrypted cloud data is developed. Original contributions are three-fold. First, relevance scores and preference factors are introduced upon keywords which enable the precise keyword search and personalized user experience. Second, a practical and very efficient multi-keyword search scheme is developed. The proposed scheme can support complicated logic search the mixed and "AND", "OR", and "NO" operations of keywords. Third, classified sub dictionaries technique is employed to achieve better efficiency on index building, trapdoor generating and query. Lastly, the security of the proposed schemes is analyzed in terms of confidentiality of documents, privacy protection of index and trapdoor, and unlinks ability of trapdoor.

**Keywords:** Cloud computing, encrypted, confidentiality, trapdoor.

## I. INTRODUCTION

The cloud computing treats computing as a utility and leases out the computing and storage capacities to the public individuals in such a framework, the individual can remotely store her date on the cloud server, namely data outsourcing, and ten make the cloud data open for public access through the cloud server. This represents a more scalable, low-cost and stable way for public data access because of the scalability and high efficiency of cloud servers, and therefore is favorable to small enterprises that the outsourced data may contain sensitive privacy information. It is often necessary to encrypt the private data before transmitting the data to the cloud servers. The data encryption, however, would significantly lower the usability of data due to the difficulty of searching over the encrypted data. Simply encrypting the data may still cause other security concerns. For instance, Google Search use SSL (Secure Sockets Layer) to encrypt the connection between search user and Google server when private data, such as documents and emails, appear in the search results. However, if the search user clicks into another website from the search results page, that website may be able to identify the search terms that the user has used. On addressing above issues, the searchable encryption has been recently developed as a fundamental approach to enable searching over encrypted cloud data, which precedes the following operations. Firstly, the data owner needs to generate several keywords according to the outsourced data. These keywords are then encrypted and stored at the cloud server. When a search user needs to access the outsourced data, it can select some relevant keywords and send the cipher text of the selected keywords to the cloud server. The cloud server then uses the cipher text to match the outsourced encrypted keywords, and lastly returns the matching results to the search user. To achieve the similar search efficiency and precision over encrypted data as that of plaintext keyword search, an extensive body of research has been developed in literature.

We introduce the relevance scores and the preference factors of keywords for searchable encryption. The relevance scores of keywords can enable more precise returned results. And the preference factors of keywords represent the importance of keywords in the search keywords set specified by search users and correspondingly enable personalized search to cater to specific user preferences. It thus further improves the search functionalities and user experience. We realize the "AND", "OR" and "NO" operations in the keyword search for searchable encryption. Compared with schemes the proposed scheme can achieve more comprehensive functionality and lower query complexity.

## II. LITERATURE SURVEY

**1) Secure ranked keyword search over encrypted cloud data**

**AUTHORS**: C.Wang, N. Cao, J. Li, K.Ren, and W. Lou

As cloud computing becomes prevalent, sensitive information are being increasingly centralized into the cloud. For the protection of data privacy, sensitive data has to be encrypted before outsourcing, which makes effective data utilization

a very challenging task. Although traditional searchable encryption schemes allow users to securely search over encrypted data through keywords, these techniques support only Boolean search, without capturing any relevance of data files. This approach suffers from two main drawbacks when directly applied in the context of cloud computing. On the on hand, users, who do not necessarily have pre-knowledge of the encrypted cloud data, have to post process every retrieved file in order to find ones most matching their interest, on the other hand, invariably retrieving all files containing the queried keyword further incurs unnecessary network traffic , which is absolutely undesirable in today's pay-as-you-use cloud paradigm. In this paper, for the first time we define and solve the problem of effective yet secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency), thus making one step closer towards practical deployment of privacy-preserving data hosting services in cloud computing.

**2) Order-preserving symmetric encryption**

**AUTHORS:** A.Boldyreva, N. Chenette, Y. Lee, and A. Oneill

We initiate the cryptographic study of order-preserving symmetric encryption (OPE), a primitive suggested in the database community by Agrawal et al. (SIGMOD '04) for allowing efficient range queries on encrypted data. Interestingly, we first show that a straight forward relaxation of standard security notions for encryption such as in distinguish ability against chosen plaintext attack (IND-CPA) is unachievable by a practical OPE scheme. Instead, we propose a security notion in the spirit of pseudorandom functions (PRFs) and related primitives asking that an OPE scheme look "as random as possible" subject to the order preserving constraint. We then deign an efficient OPE scheme and prove its security under our notion based on pseudo randomness of an underlying block cipher. Our construction is based on a natural relation we uncover between a random order preserving function and the hypergeomeytric probability distribution. In particular, it makes black box use of an efficient sampling algorithm for the latter. Cloud computing has emerging as a promising pattern for data outsourcing and high quality data services. However, concerns of sensitive information on cloud potentially causes privacy problems. Data encryption protects data security to some extent, but at the cost of compromised efficiency. Searchable symmetric encryption (SSE) allows retrieval of encrypted data over cloud.

**3 Towards preference aware relational databases**

**AUTHORS:** A.Arvanitis and G. Koutrika

In Implementing preference aware query processing, a straight forward option is to build a plug-in on top of the database engine. However, treating the DBMS as a black box affects both the expressivity and performance of queries with preferences. In this paper, we argue that preference aware query processing needs to be pushed closer to the DBMS. We present a preference aware relational data model that extends database tuples affected, their preference scores and the credibility of the preference. Our query processing strategies push preference evaluation inside the query plan and leverage its algebraic properties for finer grained query optimization. We experimentally evaluate the proposed strategies. Finally, we compare our framework to a pure plug in implementation and we show its feasibility and advantages.

**4) Preference based query personalization**

**AUTHORS:** G.Koutrika, E.Pitoura, and K. Stefanidis

In the context of database queries, computational methods for handling preferences can be broadly divided into two categories. Query personalization methods consider that user preferences are provided as a user profile separately from the query and dynamically determine how this profile will affect the query results. On the other hand, preferential query answering methods consider that preferences are explicitly expressed within queries. The focus of this chapter is on query personalization methods. We will first describe preference can be represented and stored in user profiles. Then, we will discuss how preferences are selected from a user profile and applied to a query.

## III. RELATED WORK

**DATA ENCRYPTION:**

Encryption is the process of conversion of data into a forms, referred to as a cipher textual content that cannot be effortlessly understood by means of unofficial people. Decryption is the system of changing encrypted facts back into its authentic form, so it could be understood.

Encryption is a system of changing the obvious text into cipher textual content to encrypt a report or different information in a pc way convert it into mystery code in order that it may't be used or understood until it is decoded or decrypted. You might need to encrypt a document the system will ask you to enter secter key. After that no one could be capable of recognise the data unless they have the same secret key encryption hides your data from others .that is the process of encrypting data to prevent unauthorized man or woman from viewing or editing the information

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

The primary features of facts encryption are:
1. Prevents undesirable get right of entry to to files and e mail message
2. Most powerful tiers of encryption are very hard to break.

The technique of records encryption includes sure steps the facts passes through a mathematical system called an algorithm which converts it into encrypted data called cipher textual content. These algorithms create a key after which encapsulate the message with this key .there are varieties of encryptions: asymmetric and symmetric

**3.2 Asymmetric Encryption:**

In public key (uneven) encryption, two mathematically-associated keys are used: one to encrypt the message and the other to decrypt it.

Those keys integrate to form a key pair. Asymmetric encryption presents each facts encryption and validation of the speaking parties' identities and is taken into consideration greater relaxed than symmetric encryption, however is computationally slower.

A public key encryption scheme has five most important components:

1. Plaintext: that is the textual content message to which an set of rules is applied.
2.Encryption Algorithm: It uses green seek approach to generate a secret key for changes to the plaintext
3. Public and private Keys: this is a couple of keys where one is used for encryption and the alternative for decryption.
4. Cipher text: content-that is the encrypted or scrambled message produced by making use of the set of rules to the plaintext message using key.
5. Decryption algorithm: This algorithm generates the ciphertext and the matching key to provide the plaintext.

**3.3 Symmetric Encryption**

Private Key encryption (Symmetric), also known as traditional or single-key encryption is based on secret key that is shared with the aid of each speaking events. It enquires all events which are communicating to percentage a commonplace key. The sending birthday celebration uses the secret key as a part of the mathematical operation to encrypt (or encipher) undeniable text to cipher textual content. The receiving birthday celebration makes use of the identical secret key to decrypt (or decipher) the cipher text to plain text.

Examples of symmetric encryption schemes are the RSA RC4 set of rules data Encryption standard (DES).when the usage of this shape of encryption, it is essential that the sender and receiver have a manner to trade mystery keys in a comfortable way. If a person knows the secret key and may determine out the algorithm, communications might be insecure. there may be also the need for a sturdy encryption algorithm. What this indicates is that if a person were to have a ciphertext and a corresponding plaintext message, they could be not able to determine the encryption set of rules.

## IV. THE SYSTEM ARCHITECTURE

**CLIENT:** destination IP address, source path of the file, open the file and split file transfer the file.
**SERVER:** size, on/off. Start server running and waiting to receiver receiving completed.
**DATA OWNER:**

The data owner outsources her data to the cloud for convenient and reliable data access to the corresponding search users. To protect the data privacy, the data owner encrypts the original data through symmetric encryption. To improve the search efficiency, the data owner generates some keywords for each outsourced document. The corresponding index is the create according to the keywords and a secret key. After that, the data owner sends the encrypted documents and the corresponding indexes to the cloud, and sends the symmetric key and secret key to search users.

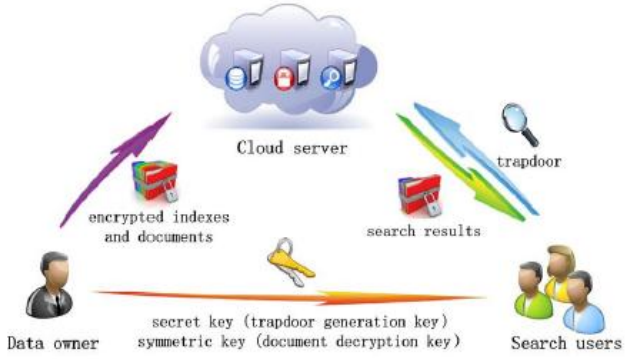


Fig 1: Representes the data transfers from one objectives to the other

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

**CLOUD SERVER:**
The cloud server is an intermediate entity which stores the encrypted documents and corresponding indexes that are received from the data owner, and provide data access and search services to search users. When a search user sends keyword trapdoor to the cloud server, it would return a collection of matching documents based on certain operations.

**SEARCH USER:**
A search user queries the outsourced documents from the cloud server with following three steps.
First: the search user receives both the secret key and symmetric key from the data owner.
Second: according to the search keywords, the search user uses the secret key to generate trapdoor and sends it to the cloud server. Last, she receives the matching document collection from the cloud server and decrypts them with the symmetric key.

**MULTIKEYWORD SEARCH**:
Data outsourcing is the major challenge in the cloud computing where the data is stored in the secured format of encryption and then retrieving the data from the cloud storage with the relevant information is a huge task since it requires the accurate data retrieval system so we propose a new approach of nearest neighborhood query retrieval process with the key word searching technique on secured cloud data so we use the process of the key management where the user and owner have the full trust and the data is more secured transposed with the authorization of security with the transpose key to encrypt and decrypt the data so we implement a scheme for efficient data retrieval using AES algorithm, triple DES, RC4, RSA algorithms are used for key word based.

## V. CONCLUSION

Investigated on the "An efficient search technique from document retrieval from encrypted cloud" data, and proposed two FMS schemes. The FMS I includes both the relevance scores and the preference factors of keywords to enhance more precise search and better users' experience, respectively. Following search technique achieves secure and efficient search with practical functionality, i.e., "AND", "OR" and "NO" operations of keywords. Furthermore, we have proposed the enhanced schemes supporting to improve efficiency. For the future work, we intend to further extend the proposal to consider the extensibility of the file set and the multi-user cloud environments. Towards this direction, we have made some preliminary results on the extensibility and the multiuser cloud environments. Another interesting topic is to develop the highly scalable searchable encryption to enable efficient search on large practical databases.

## REFERENCES

[1] H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, "An smdpbased service model for interdomain resource allocation in mobile cloud networks," IEEE Transactions on Vehicular Technology, vol. 61, no. 5, pp. 2222–2232, 2012.
[2] M. M. Mahmoud and X. Shen, "A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 10, pp. 1805–1818, 2012.
[3] Q. Shen, X. Liang, X. Shen, X. Lin, and H. Luo, "Exploiting geodistributed clouds for e-health monitoring system with minimum service delay and privacy preservation," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 2, pp. 430–439, 2014.
[4] T. Jung, X. Mao, X. Li, S.-J. Tang, W. Gong, and L. Zhang, "Privacypreserving data aggregation without secure channel: multivariate polynomial evaluation," in Proceedings of INFOCOM. IEEE, 2013, pp. 2634–2642.
[5] Y. Yang, H. Li, W. Liu, H. Yang, and M. Wen, "Secure dynamic searchable symmetric encryption with constant document update cost," in Proceedings of GLOBCOM. IEEE, 2014, to appear.
[6] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multikeyword ranked search over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, 2014.