A Secured and Searchable Encryption Algorithm for Cloud Storage

Krati Mehto
CSE Department
AITR, Indore [M.P]
India

Rahul Moriwal
CSE Department
AITR, Indore [M.P]
India

ABSTRACT

Cloud computing is a new generation technology which efficiently support the client oriented services. Now in these days there are a number of applications which consumes the cloud storage service for storing and retrieving information. In such conditions the data owner management and privacy preservation cryptographic techniques are utilized frequently. But due to cryptographic technique of security implementation the data leave their own format and converted into other unreadable format. Due to this retrieval of required information becomes complex. Therefore in this paper a proposed solution incorporate the hash table management and indexing techniques to keep track the actual data contents in terms of document features which may help for encrypting user data and identifying the user data and privacy.

Keywords

Cloud computing, MD5, AES, Cloud storage.

1. INTRODUCTION

The Internet access becomes available in the recent years, Cloud computing is an internet based technology; it is using hardware and software as computing resources to provide service through internet. Cloud computing is being widely used now-a-days enabling the end user to create and use software from anywhere at any time without worrying about the execution of the technical information.

Cloud computing technology provide unlimited resources and services like data storage service which helps to manage the user data. Now a days, with the help of dynamic data operation with computation user can store their data in cloud. Which makes the copy of data for further updating and verification of the data loss. Here with the help of cryptographic technique data can be secured from unauthorized user or access.

The benefits of the cloud storage are flexible with reduced cost and they also manage the data loss risk and so on.[1]

1.1 Cryptography

Cryptography is the science and study of secret writing, whereby plaintext (or clear text) is transformed into cipher text (sometimes called a cryptogram). The whole process of converting plaintext into cipher text is called encipherment or encryption; the reverse process is called decipherment or decryption. Both processes are controlled by a cryptographic key.The branch of both cryptography and cryptanalysis is called cryptology.

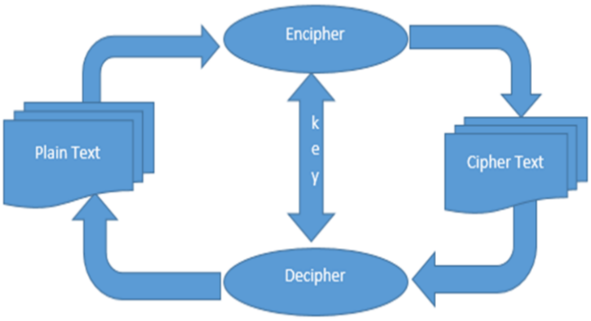

Fig 1: Secret writing.

1.2 Data Security

Classical cryptography provided secrecy for information sent over channels where eavesdropping and message interception was possible. The sender selected a cipher and key for encryption and either gave it directly to the receiver or else sent it indirectly over a slow but secure channel (typically a trusted courier). Messages and replies were transmitted over the insecure channel in cipher text. [2]

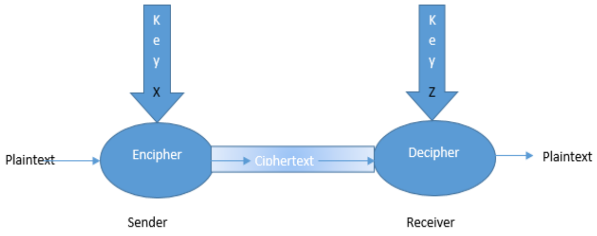

Fig 2: Classical information channels

To provide an information retrieval function while addressing the security and privacy issues, the concept of searchable encryption was introduced in an earlier study [3]. With searchable encryption, the entity performing the retrieval service is not allowed to learn the content of the queries and responses. Instead, some additional encrypted index terms (serving as keywords or hints) are used for the data search process. The entity uses a cryptographic algorithm similar to decryption for finding the correspondence between the encrypted query and the encrypted data content. They summarize the essential requirements specified in [3] for searchable encryption in terms of privacy and security issues from the view point of unqualified users as follows:

Privacy of data: No one can uncover information about data content from the query and response as well as the cipher text itself.

Privacy of the data owner: No one can learn about the identity of the data owner from the encrypted content.

Privacy of the retriever: No one can learn the identities of the intended retrievers for the encrypted content.

Existing searchable encryption schemes [4–6] satisfying the above requirements involve critical limitations in terms of searching efficiency. They require a preliminary complex cryptographic manipulation before the keyword matching verification of the index terms in the query and encrypted content. Since the searching entity should perform the above computation for every item of encrypted contents exhaustively per search request, the retrieval performance declines significantly in cloud information retrieval systems in which frequent queries are sent to the CSP. This problem becomes more serious in which there are numbers of intended receivers for one instance of data content from the same owner, resulting in redundant instances of encrypted contents owing to unique identity of each receiver. It is also important to note that searchable encryption schemes [3] were originally introduced for keyword-based searches of encrypted data which are suitable for a secure one-to-one communication. Otherwise, the key-generating authority should distribute two types of private keys to intended users, thus doubling the key storage requirements: one for access to encrypted content and one for the actual decryption of the encrypted content. Because searchable encryption schemes using attribute-based encryption (ABE) such as Hidden Vector Encryption (HVE) [4] and Inner Product Encryption (IPE) [5, 6] follow the same assumptions, the one-to-many communication property of ABE is restricted in these variations.

In [10], they propose a new searchable encryption scheme that exploits ABE with scrambled attributes to handle the problems described above, specifically, the presence of redundant encrypted data for the same message, poor expressiveness regarding access policy, and the concentration of computational overhead on the searching entity. In ABE, the access policy can be represented as Boolean expressions which consist of logical operators such as AND or OR with various attributes describing who is eligible to access the data content. Under this approach, the data owner can specify both of fine-grained access policy and searching keyword set which is required to retrieve its data under the access policy. To retrieve the encrypted content in cloud storage, the retriever makes index terms from its private key satisfying the access policy made up of keywords associated with the content, where these index terms are only used for data accessing in the cloud storage system. Therefore, the CSP cannot learn which keywords are associated to the retriever's query. Thus, our scheme is suitable for one-to-many content distribution without a sacrifice of the nature of ABE. Considering that a storage service may be operated in a one-upload-many-download manner, this advantage is an important feature. In addition, this concept might also have something in common with intrusion tolerant systems (ITSs) [22, 23] in which data leakage is possible during exposure time of virtual machines allowing an adversary to get sensitive data from information systems. They present their contributions below:

1. Their searchable encryption scheme provides rich expressiveness of index terms by exploiting ABE, thus providing data security.

2. Their scheme enhances the searching efficiency by balancing operational overhead among the CSP and other users involved in the cloud storage-based information retrieval service. Through simple comparisons on index terms common to all users, instead of an exhaustive search with cryptographic calculations, the searching process becomes more compatible with existing database management system (DBMS) mechanisms.

3. Their scheme is more suitable for a cloud storage service which operates in one-upload-many-download manner due to its use of ABE (which already provides secure one-to-many communications.)

1.3 Related work

Attribute-Based Encryption (ABE):
Attribute-based encryption was firstly introduced by Sahai and Waters [12], in which they suggested that one's identity can be viewed as a combination of several attributes expressing the characteristics of the user in the form of access policy by using Boolean expressions such as AND, OR, or NOT. Later studies are broadly categorized into key-policy attribute-based encryption (KP-ABE) [5, 12, 13] and cipher text-policy attribute-based encryption (CP-ABE) [14–16] studies. In KP-ABE, the access policy is associated with keys corresponding to attributes implying that an encrypt or is not authorized to grant access to the encrypted content except of descriptive attributes for the data by the encryptor's choice. On the other hand, CP-ABE is complementary to KP-ABE by enabling encrypt or to specify access policy combined with the cipher text. Both schemes allow secure one-to-many communications such as targeted broadcasts for a specific group and individual user according to their attributes, some studies [6–9] suggested modification of ABE schemes by hiding the access policy. These schemes operate on the assumption that the data owner directly delivers the cipher text to the receiver without an intermediate third party. In other words, when adopting those approaches directly in cloud storage, decryption keys can be exposed to an unauthorized third party. Hence, they are not feasible for data retrieval services in the cloud storage systems because the test procedure allows the CSP to learn which attributes the user has.

2. BACKGROUND

2.1 Modern Cryptography:

The modern cryptography can be divided into many fields. Major one are discussed here.

2.1.1 Symmetric-key cryptography:

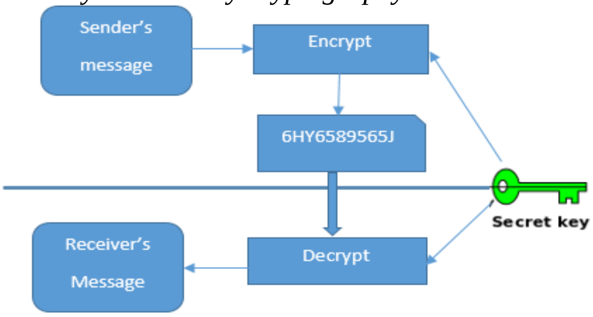

Fig 3: secret key encryption

Symmetric-key cryptography, where only one key is used for encryption and decryption also. So here both the sender and receiver use the same key.

2.1.2 Public-key cryptography:

Public-key cryptography, where two different keys are used to encrypt and decrypt the data. So one key is used for encryption and other one is used for decryption process, but in symmetric-key system the same key is used for encryption and decryption of a message.

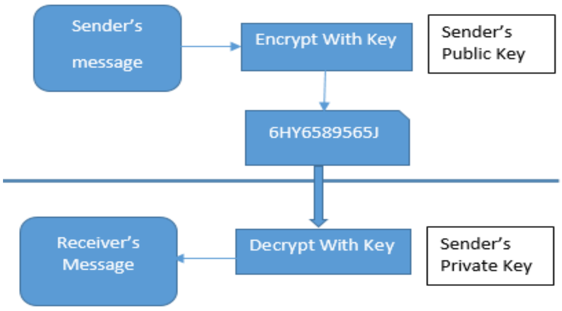

Fig 4: public key encryption

So disadvantage of this system is the key management necessary to use them securely. *[11]*

2.2. Searchable Encryption

The original goal of searchable encryption is to provide privacy-preserving keyword searches of encrypted data against an intermediate gateway such as a mail server or a network router, which involves a message exchange process between the sender and the receiver. The first searchable encryption scheme was the Public-key Encryption with Keyword Search (PEKS) scheme based on Identity-Based Encryption (IBE), originally proposed by Boneh et al [3]. Since PEKS is devised to forward the encrypted contents to a designated receiver with its unique identity, this scheme restricts expressiveness as regards access policy. To provide better expressiveness, other searchable encryption schemes based on ABE are introduced. One of representative works is Hidden Vector Encryption (HVE) [4]. HVE is more advanced work compared to PEKS, as it provides Conjunctive and range queries. PEKS and HVE schemes are not suitable for one-upload-many-download cloud storage systems because they mainly focus on one-time message-delivery scenarios. [10]

3. LITERATURE SURVEY

wang et al[17] motivate and solve the problem of supporting effective ranked keyword search for providing efficient use of remotely stored encrypted data in Cloud. They firstly give a fundamental scheme and show that by same existing searchable encryption method, it is not efficient to achieve ranked search. So they appropriately weaken the security guarantee, to solve this security problem they developed cryptography first OPSE, and derive an efficient one-to-many order-preserving mapping function, which allows the effective RSSE to be designed. Through thorough security analysis, they show that their proposed solution is secure and maintain the privacy, while correctly realizing the aim of ranked keyword search. And also shows that their solution enjoys “as-strong-as possible” security guarantee compared to conventional SSE schemes, Note that in their design, they focus on single keyword search.

Advantage- Their solution is secure and privacy-preserving, and provide “as-strong-as possible” security.

Disadvantage- Their solution is only efficient for single keyword search.

Jyunet al[18]introduce a strong and searchable encryption scheme for outsourcing the data in cloud by considering both type of data i.e. numeric and non-numeric. In other aspect, their scheme provides some fault-tolerance for cloud computing.

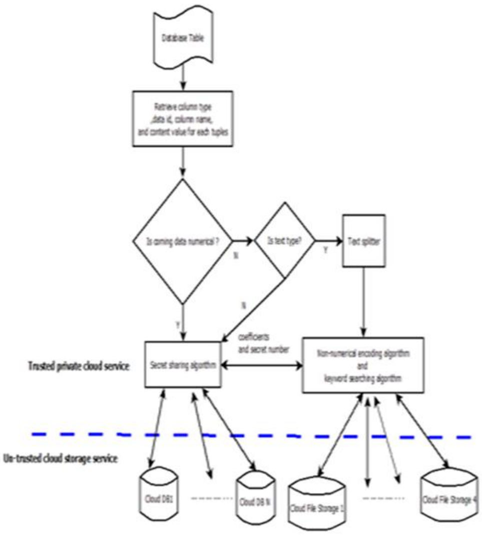
Database Table
Retrieve column type ,data id, column name, and content value for each tuples
Is coming data numerical ?
N
Is text type?
Y
Text splitter
N
Y
coefficients and secret number
Trusted private cloud service
Secret sharing algorithm
Non-numerical encoding algorithm and keyword searching algorithm
Un-trusted cloud storage service
Cloud DB1
Cloud DB N
Cloud File Storage 1
Cloud File Storage 4

Fig 5: System architecture [18]

The design of this scheme is shown in Figure. The architecture is partitioned into two parts: trusted private cloud service and un-trusted cloud service. The un-trusted cloud service provides several cloud databases (noted by CDB) and cloud file storages (noted by CFS). Cloud database service and cloud file storage service may be provided by different cloud service providers. In their scheme, there are N cloud databases and four cloud file Storages. Note, each CDB and CFS has its unique id. Also note CDBi as the CDB with id i and CFSi as CFS with id i, respectively. And finally they analyze their proposed scheme with three aspects: security analysis, performance analysis, and fault tolerance.

Advantage- They provide a scheme for data outsourcing in cloud for both numeric and non-numeric data.

Disadvantage-The computation cost of this scheme is high.

Ming et al [19] identify the requirements of system and challenges to achieve privacy for searchable cloud data services for outsourcing. They present a general methodology for using searchable encryption techniques, which allow data security without any leaking of data. Here they discuss three usable search operation functionalities: similarity search, supporting result ranking, and search over structured data. In some scenarios, the data owner and user can be the same person. But they consider the cloud server to be semi-trusted. Thus, the CSP and owners of data are not assumed to be in the same trust domain. In addition, some other users may also try to access the data beyond their privileges.

Advantage-They introduce a scheme where they prevent from leaking information about search data and user query.

Disadvantage- Due to symmetric-key-based solutions more complex data utilization cannot be efficiently evaluated.

Wang et al [20] introduces, that conventional data utilization services basically based on plaintext keyword search. So downloading all the data and decrypting locally is clearly impractical, due to the large amount of bandwidth cost in cloud scale systems. In this paper, they define and solve the challenging problem very first time of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and also establish a strict set of privacy requirements for utilization of secure cloud data system to become a reality. Among various semantics of multi-keyword, they choose the efficient principle of “coordinate matching” which means as many matches as possible for finding the similarity between data document and search query, and further use “inner product similarity” for similarity measurement.

Advantage- They provide strict privacy to make cloud data accessing more secure with multi-keyword ranked search.
Disadvantage- Their solution does not tolerate keyword spelling error.

Wang et al [21] proposed a multi keyword fuzzy search scheme by taking advantages of locality-sensitive hashing technique. Their proposed scheme attain fuzzy matching through algorithmic design without extending the index file. Here they remove the need of predefined dictionary and without increasing the search complexity it effectively supports multiple keyword fuzzy search. Analysis on very large scale and many experiments on real-world data shows that their proposed scheme is very secure and efficient. Existing solutions for multi keyword exact search does not permit keyword spelling error. The current fuzzy search schemes expanded index that include all possible misspelling keyword, which increase index file size and search complexity. They achieve this by several novel designs based on locality-sensitive hashing (LSH) and Bloom filters. Extensive analysis shows that their scheme is secure, efficient and accurate.

Advantage- They provide fuzzy matching without any search complexity.

4. PROPOSED WORK

This section includes involved work and identified issues in system in addition to that an optimum solution is also provided. The cloud environment provides support for efficient computing and enables to provide the storage solutions at the remote end. The main aim is to address the following issues in the existing cloud storage:

1. ***Data security***: the data is placed on the cloud which is not much secured due to third party access and treads therefore the data security in cloud storage is required
2. ***Data owner and client privacy management:*** the data owner and client in not distinguishable using the data additionally the privacy on such data is access is required.
3. ***Searchable data space***: the cryptographic manner of data security converts the formats and not a bit of data recovered during the information retrieval.

We have discussed above, about the issues and challenges now we will provide the solution steps that are described below. In order to provide end to end solution for the cloud storage the following solution steps are included.

1. **Authentication management:** in authentication management the system and user attributes are recovered additionally the one time password is included to manage the secure authentication.
2. **Cryptographic data security:** in this phase the MD5 and AES based hybrid cryptographic algorithm is consumed for providing the security.
3. **Providing the search solution over the encrypted data:** the keyword based search system is provided for identifying the user and their data during different data retrieval operations.

The proposed scheme can be understood using the given figure 5.According to the given figure the proposed security technique involve the following steps of authentication and data preserving technique.

1. Client node is an end client system who wants to store or retrieve the data from the secure server. In this step the end client initiate the authentication by making the request from the server.
2. In this phase the server system trigger the authentication server for finding the user credentials and data attributes and ask for the security questions, in this phase the OTP is applied to make secure the communication between client and server.
3. After authenticating the user access the system ask for user id, password and OTP again here the OTP works as the salt for the encryption and validation.
4. In this step user initiate the communication and data request from the server, during this the MD5 and AES algorithm is organized for encrypting the data additionally the following information is preserved into the attribute MAP.

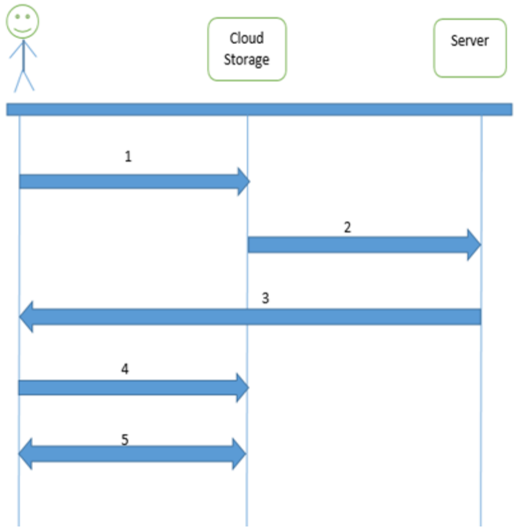

Fig 6: security management

The below given working steps can summarized as:

1. User ID-Id is used to identify user uniquely. It is provided by server to user when they interact with server first time. And after that when user wants to access the system.
2. Password-Password is a secret key or we can say which is used by user at the time of login. Every user has its own password.
3. Session key-Session key define the particular session or time, as much time user connected with server.

4. Text file features as frequent token.
5. Original file name-OFN is the real name of file by which it is accessed.(stored or retrieved)
6. Mapped file name- MFN is that name by which OFN is mapped and then stored in hash table. So that our original file could remain save.
5. In final step user can access the information and data using the server in this step the KNN algorithm is applied over MAP data for finding the user targeted information from search space.

5. CONCLUSION&FUTUREWORK

In this paper we discussed various method of searchable encryption to secure the data in cloud storage. Also we discussed about cryptography methods which helps to convert the data from readable to unreadable form so our data could be saved in cloud storage from adversary. By studying all these paper we can conclude that the essence of security for cloud storage is very necessary so that client could feel secure while accessing the cloud storage services. In this paper we proposed a scheme for secure data accessing with maintaining its privacy by using strong cryptographic algorithm. Our future work will attempt to enhance the feasible solution.

6. REFERENCES

[1] SwapnilV.Khedkar et al, Data Partitioning Technique toImprove Cloud Data Storage Security. / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3347-3350

[2] http://faculty.nps.edu/dedennin/publications/Denning-CryptographyDataSecurity.pdf.

[3] Boneh D, Di Crescenzo G, Ostrovsky R, Persiano G. Public key encryption with keyword search. In: Cachin C, Camenisch J, editors. Advances inCryptology, EUROCRYPT 2004, vol. 3027. Berlin/Heidelberg: Springer; 2004. p. 506–22.

[4] Boneh D, Waters B. Conjunctive, subset, and range queries on encrypted data. In: Vadhan S, editor. Theory of cryptography, vol. 4392. Berlin/ Heidelberg: Springer; 2007. p. 535–54.

[5] Katz J, Sahai A, Waters B. Predicate encryption supporting disjunctions, polynomial equations, and inner products. 27th annual international conference on Advances in cryptology, EUROCRYPT'08. Berlin, Heidelberg: Springer-Verlag; 2008. p. 146–62.

[6] Lewko A, Okamoto T, Sahai A, Takashima K, Waters B. Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption. Advances in Cryptology, EUROCRYPT 2010, vol. 6110. Springer; 2010. p. 62–91.

[7] Kapadia A, Tsang PP, Smith SW. Attribute-based publishing with hidden credentials and hidden policies. In: The 14th annual Network and Distributed System Security Symposium (NDSS 07) to appear; 2007.p.179–92.

[8] Yu S, Ren K, Lou W. Attribute-based content distribution with hidden policy. In: 4th Workshop on secure network protocols, 2008. NPSection 2008; 2008. p. 39–44.

[9] Nishide T, Yoneyama K, Ohta K. Attribute-based encryption with partially hidden encryptor-specified access structures. Applied cryptography and network security, vol. 5037. Berlin/Heidelberg: Springer; 2008. p. 111–29.

[10] Dongyoung Koo, JunbeomHur, Hyunsoo Yoon, "Secure and efficient data retrieval over encrypted data usingattribute-based encryption in cloud storage",2012 Elsevier Ltd. All rights reserved.

[11] CryptographyFrom Wikipedia, link: Cryptography%20-%20Wikipedia,%20the%20free%20encyclopedia.html,the free encyclopedia.

[12] Sahai A, Waters B. Fuzzy identity-based encryption. In: Cramer R, editor. Advances in cryptology, EUROCRYPT 2005, vol. 3494. Berlin/Heidelberg: Springer; 2005. p. 557–57.

[13] Goyal V, Pandey O, Sahai A, Waters B. Attribute-based encryption for fine-grained access control of encrypted data. In: Proceedings of the 13th ACM conference on Computer and communications security, CCS '06. New York, NY, USA: ACM; 2006. p. 89–98.

[14] Bethencourt J, Sahai A, Waters B. Ciphertext-policy attribute-based encryption. In: IEEE symposium on security and privacy; 2007. p. 321–34.

[15] Waters B. Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization. In: Catalano D, Fazio N, Gennaro R, Nicolosi A, editors. Public Key Cryptography, PKC 2011, vol. 6571. Berlin/Heidelberg: Springer; 2011. p. 53–70.

[16] Cheung L, Newport C. Provably secure ciphertext policy abe. In: Proceedings of the 14th ACM conference on Computer and communications security, CCS '07. New York, NY, USA: ACM; 2007. p. 456–65.

[17] Wang, li et al [2010]- Secure Ranked Keyword Search over Encrypted Cloud Data

[18] Jyun-Yao et al [2012] - A Searchable Encryption Scheme for Outsourcing Cloud Storage

[19] Ming li, yu et al [2013] - Toward Privacy-Assured and Searchable Cloud Data Storage Services

[20] Wang, li et al [2014] - Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data

[21] Wang, Yu et al [2014] Privacy-Preserving Multi-Keyword Fuzzy Search over Encrypted Data in the Cloud

[22] Saidane A, Nicomette V, Deswarte Y. The design of a generic intrusion-tolerant architecture for web servers. IEEE Trans Dependable Secure Comput 2009;6(1):45–58.

[23] Sousa P, Bessani AN, Correia M, Neves NF, Verissimo P. Highly available intrusion–tolerant services with proactive-reactive recovery. IEEE Trans Parallel Distrib Syst 2010;21(4):452–65.