Privacy Preserving Multi-keyword Search Encrypted Using Cloud Data Storage

Ms. Apurva S. Solanke (*Author*)
MSS College of Engineering & Technology ,
Jalna , India

Abstract— Public cloud is now very quick growing trend for storing user's data. Most of the users now a day's placing their personal and professional data on the cloud. Cloud computing has won over a lot of interest from different areas since it provides Systematic resource management, Economical cost , Fast deployment, scalability, availability, low cost service over traditional storage solutions. The imaginative in cloud computing has motivated the data owner to deploy their data from local sites to profitable public cloud for enormous elasticity and profitable savings. At the same time secrecy of remotely stored data on untrusted cloud server is big responsibility. In order to diminish these responsibility of sensitive data such as E-mails, reports, personal information are outsourced in encrypted form using encoding system. The security concerns in cloud computing motivate the study on secure keyword search. The search techniques which are used on plain text cannot be used over encrypted data. The existing solutions supports only identical keyword search, semantic search is not supported. As we don't want to disclose neither keyword from query nor query pattern, we have developed fully privacy preserving system by encrypting search pattern as well as secret key. Indexing has been developed to build an index of keywords from documents. Index will be used to retrieve documents in response to search query by using the principle of keyword matching. This paper has analyzed and implemented Lucene indexing algorithm. Ranking of the results has been developed so as to improve the search result correctness as well as to enrich the user searching experience.

Key Words : Multi-Keyword search, Coordinate Matching, Keywords, Index Generation, Trapdoor

I. INTRODUCTION

Public cloud is now very quick growing trend for storing user's data. Most of the users now a day's placing their personal and professional data on the cloud. Cloud computing has won over a lot of interest from different areas since it provides Systematic resource management, Economical cost , Fast deployment, scalability, availability, low cost service over traditional storage solutions. Cloud Computing provides us a means by which we can access the applications as utilities, over the Internet. It allows us to create, configure, and customize applications online. Cloud computing are new type of computing paradigm which enables sharing of computing resources over the internet. The cloud characteristics are on-demand self-service, location independent network access, ubiquitous network access and usage based pay. Due to this charming features private and public organization are deploy their large amount of data on cloud storage. A semantic secure multi-keyword search scheme over encrypted cloud data is proposed in this paper. The semantic search is not only support exact keyword matched or structure matched but also supports the real intent of user search. The relevance score between documents and query keywords is calculated and files are returned in ranked order.

It allows us to create, configure & customize application online. The term cloud refers to network or internet. In other word, we can say that cloud is something which is present at remote location.

The search facility and privacy protective over encrypted cloud data are essential. If we study huge amount of data documents and data users in the cloud, it is hard for the necessities of performance, usability, plus scalability. Concerning to encounter the real data recovery, the huge amount of data documents in the cloud server achieve to outcome relevant rank instead of returning undistinguishable outcomes. Ranking scheme cares multiple keyword search to recover the search correctness. Cloud provides services over network i.e. on public network or private network i.e. WAN, LAN, or VPN. Today's Google network search devices, data users offer set of keywords instead of unique keyword search importance to retrieve the maximum significant data. Coordinate matching is a synchronize pairing of query keywords which are relevance to that document to the query. cloud computing refers to manipulating, referring & accessing the application online. It offers online data storage, infrastructure & application.

Cloud computing is come paradigm where large pool of system are connected in private or public network to provide dynamically scalable infrastructure for application data & file storage.

Due to inherence safety and privacy, it remains the interesting job on behalf of how to relate the encrypted cloud search. The difficult of multi-keyword ranked search over encrypted cloud data is resolved by using stringent privacy necessities then numerous multi-keyword semantics. Among numerous multi-keyword ranked semantics, we choose coordinate matching. Our contributions are summarized as follows,

1) For the first time, we explore the problem of multi keyword ranked search over encrypted cloud data, and establish a set of strict privacy requirements for such a secure cloud data utilization system.

2) We propose two MRSE schemes based on the similarity measure of "coordinate matching" while meeting different privacy requirements in two different threat models.

3) Thorough analysis investigating privacy and efficiency guarantees of the proposed schemes is given, an experiments on the real-world dataset further show the Proposed schemes indeed introduce low overhead on computation and communication.

The valuable data such as social security number, email, Personal health records and organizations financial information must be stored securely. Solution is encryption of data at client side before outsourcing. But if you encrypt data the searching over chipper text is challenging. The existing search techniques are only applied on plain text data. The trivial solution of downloading all the data and decrypting locally is clearly impractical due huge amount of bandwidth cost in cloud scale system. Searchable encryption allows storing data in encrypted format and you can apply keyword search over chipper text data.

Due to this charming features private and public organization are outsourcing their large amount of data on cloud storage. Organization can purchase only needed amount of storage from CSP to fulfill their data storage need instead of maintaining their own data storage. The data owner is relieved from purchasing hardware and software to manage data themselves. Instead of these tremendous advantages.

Cloud computing transforms the way information technology (IT) is expended and oversaw, promising enhanced expense efficiencies, quickened development, speedier time-to-market, and the capacity to scale applications on interest (Leighton, 2009).[1] As per Gartner, while the buildup developed exponentially amid 2008 and proceeded since, it is clear that there is a noteworthy movement towards the cloud computing model and that the advantages may be significant (Gartner Hype-Cycle, 2012). Be that as it may, as the cloud's state processing is rising and growing quickly both theoretically and actually, the legitimate/contractual, monetary, administration quality, inter-operability, security and protection issues still posture critical difficulties. In this part, we depict different services and organization models of distributed computing and recognize significant difficulties.

We consider the issue of building a safe cloud storage services on top of an open cloud foundation where the service 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 encrypted data without the decoding key – i.e., given encryptions E(m1),...,E(mt) of m1,...,mt, one can efficiently process a smaller cipher text that encrypts f(m1,...,mt) for any efficiently calculable capacity f. This issue was postured by Rivest et al. in 1978. [3]Completely homo morphic encryption has various applications. For instance, it empowers private queries to a search engine– the user presents an encrypted query and the search engine processes a brief encrypted answer while never taking a gander at the query in the clear. It likewise empowers looking on encrypted data – a user stores encrypted files on a remote file server and can later have the server recover just files that (when decoded) fulfill some boolean limitation, despite the fact that the server can't unscramble the files all alone. All the more comprehensively, completely homo morphic encryption enhances the efficiency of secure m.

We concentrate on the issue of looking on data that is encrypted using a public key system. [5] Consider user Bob who sends email to user Alice encrypted under Alice's public key. An email portal needs to test whether the email contains the keyword "urgent" so that it could course the email in like manner. Alice, then again does not wish to give the entryway the capacity to decrypt every one of her messages. We build a component that empowers Alice to give a key to the passage that empowers the entryway to test whether "urgent" is a keyword in the email without learning whatever else about the email. We allude to this component as Public Key Encryption with keyword Search

PROPOSED SYSTEM:

A data hosting service in the cloud that involves three different entities, the owner of the data, the user of the data and the server of the cloud. The owner of the data first registers in the cloud using cloud computing services. The owner of the data has a collection of F data documents to be outsourced to the server in the encrypted C form. To enable search capability on C for effective data utilization, the data owner will first build a search index I using F's Lucene Indexer before outsourcing, and then outsource both the index I and the collection of encrypted documents C to the cloud server. The work deals with efficient algorithms to assign identifiers (ID) to users in the cloud in such a way that the FILE identifiers are anonymous using a distributed calculation without central authority as the data is encrypted.

Since there are N nodes, this assignment is essentially a permutation of the integers {1 ... .N} with each FILE that is known only by the node to which it is assigned. Our main algorithm is based on a method of anonymously sharing simple data and results in methods for the efficient exchange of complex data. To search the collection of documents for certain keywords, an authorized user who has an identification and a specific designation acquires a corresponding K through our search control mechanisms. Upon receiving T from a data user, the server in the cloud is responsible for searching the index I and then returns the corresponding set of encrypted documents. To improve the accuracy of document retrieval, the cloud server must classify the search result according to some classification criteria (for example, coordinate match) and assign anonymous FILE ID [6] to the user in the cloud to Make the data cloud more secure. In addition, to reduce the cost of communication, the user of the data can send an optional k number together with the trap door T, so that the server in the cloud only sends the top-k documents that are most relevant to the query of search.

OBJECTIVE :

The new scheme must be built in such a way that any authorized users can do a search on encrypted data on• multiple keywords. The new scheme must facilitate users who can query the database provided that they possess so called trapdoors for• the search terms that authorize the end users to include them in their queries. The new scheme must offer multiple keyword searches in a single query and ranks the results so the end user can• retrieve the most relevant matches in an ordered manner. The new scheme must provide permissions to only authenticated owners to outsource the data to the cloud [5].•

ENCRYPTION ALGORITHM:

Blowfish is a popular security algorithm that was developed by Bruce Schneier in the advent of the year 1994. The algorithm works on the same line as DES and consumes block blocks with blocks of a size of 64 bits. Blowfish became quite popular after its arrival, just because Bruce Schneier [1] himself is one of the most famous experts in cryptology and, above all, the algorithm is not patented, open source is free and available for its use and modifications. Blowfish is a 64-bit block cipher with a variable length key. Define 2 different boxes: S boxes, one box P and four boxes S [3]. Taking into account that P box P is a one-dimensional field with 18 values of 32 bits. The tables contain variable values; those can be implemented in the code or generated during each initialization. The frames S S1, S2, S3 and S4 each contain 256 32-bit values. Blowfish is a symmetric encryption algorithm, which means that it uses the same secret key to encode and decrypt messages. Blowfish is also a block cipher [5], which means that it divides the message into

blocks of fixed length during encryption and decryption. The block length for Blowfish is 64 bits; Messages that do not have a size of multiples of eight bytes must be filled. Blowfish consists of two parts: key expansion and data encryption. During the expansion stage of the key, the key entered becomes several matrices of sub-keys in a total of 4168 bytes. There is the matrix P, which is eighteen boxes of 32 bits, and the boxes S, which are four matrices of 32 bits with 256 entries each. After initialization of the string, the first 32 bits of the key are XORed with P1 (the first 32-bit box in the matrix P). The second 32 bits of the key are XORed with P2, and so on, until all 448 or fewer key bits have been XORed. Cycle through the key bits returning to the beginning of the key, until the entire set P has been processed. XORed with the key. Encrypt the zero string with the Blowfish algorithm, using the modified P matrix above, to get a block 64 bits. Replace P1 with the first 32 output bits, and P2 with the second 32 output bits (from the 64-bit block). Use the 64-bit output as input again in the Blowfish encryption, to get a new block of 64 bits. Replace the following values in the matrix P with the block. Repeat for all the values in the matrix P and all the squares S in order.Encrypt the whole zero chain using the Blowfish algorithm [12], using the modified P matrix above, to obtain a block of 64 bits. Replace P1 with the first 32 output bits and P2 with the second 32 output bits (from the 64-bit block). Use the 64-bit output as input again in the Blowfish encryption, to get a new block of 64 bits. Replace the following values in the matrix P with the block. Repeat for all the values in the matrix P and all the squares S in order.

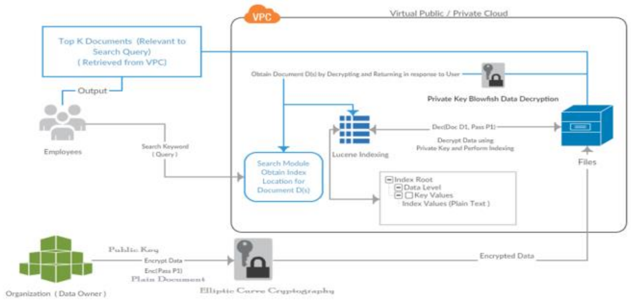
MODIFIED ALGORITHM :

This system basically uses the Blowfish encryption algorithm [12] to encrypt the data file. This algorithm is a 64-bit block cipher with a variable length key. This algorithm has been used because it requires less memory. It uses only simple operations, therefore, it is easy to implement. It is a 64-bit block cipher and is a fast algorithm for encrypting data. It requires a 32-bit microprocessor at a rate of one byte for every 26 clock cycles. It is a variable length key block encryption of up to 448 bits. Blowfish contains 16 rounds. Each round consists of XOR operation and a function. Each round consists of key expansion and data encryption. The key expansion generally used to generate initial contents of a matrix and the data encryption uses a network of 16 round of Feistal [14]. Simple text and key are the entries of this algorithm. 64 bit Normal text is taken and divided into two 32-bit data and in each round the given key is expanded and stored in 18 p-array and gives 32bit key as input and XORed with previous round data. The functionality consists in dividing a 32-bit input into four bytes and using them as indexes in an S matrix. Search results are aggregated and XOR together to produce the result. In round 16 there is no function. The output of this algorithm must be 64-bit encrypted text. It is having a function to iterate 16 times of network. Each round consists of a permutation dependent on the key and a key and a substitution dependent on the data. All operations are XOR and additions in 32-bit words. The only additional operations are four index data search tables indexed for each round.

MULTI-KEYWORD SEARCH :

Multi-keyword positioned collogue enable exact, victorious and assured inquiry over scrabbled adaptable cloud data. Security examination had displayed that different multi-keyword seek design may do arrangement of reports and record, trapdoor assurance, trapdoor unlinkability, and covering access case of the request customer in a simple way. Inside this structure, we utilize a successful record to furthermore upgrade the interest adequacy, and get the outwardly disabled limit system to mask get to case of the chase customer. This structure developed the accessible encryption for multi-watchword situated investigate the limit data. Specifically, by considering the broad number of outsourced documents (data) in the cloud and utilized the significance score and k-nearest neighbor methodologies to develop a capable multi-catchphrase look for plot that can reestablish the situated inquiry things in light of the precision.

ARCHITECTURE:


VPC
Virtual Public / Private Cloud
Top K Documents (Relevant to Search Query) (Retrieved from VPC)
Output
Employees
Search Keyword (Query)
Obtain Document D(s) by Decrypting and Returning in response to User
Private Key Blowfish Data Decryption
Search Module Obtain Index Location for Document D(s)
Lucene Indexing
Dec(Doc D1, Pass P1)
Decrypt Data using Private Key and Perform Indexing
Files
Index Root
Data Level
Key Values
Index Values (Plain Text)
Public Key
Encrypt Data
Enc(Pass P1)
Plain Document
Elliptic Curve Cryptography
Encrypted Data
Organization (Data Owner)

1. Cloud server has private key. Private Key will be used for Blowfish decryption

2. Each Cloud user will also have private key. It will be used for Encryption.

3. User wants to store a document on cloud.

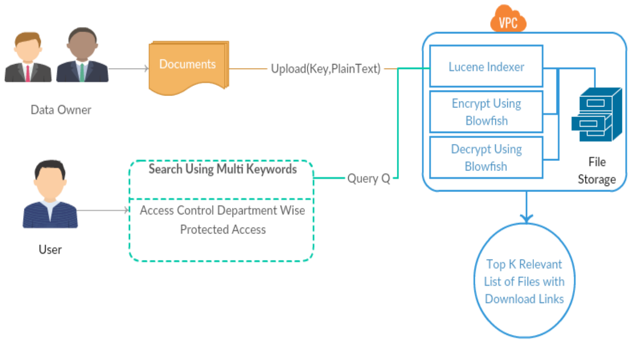
4. First he will encrypt secret key (e.g. password) and Document. Then he will upload the encrypted document and encrypted secret key

5. Server will generate index for the new document by firstly decrypting the document.

6. After creating index, secret key will get decrypted by server using Blowfish with private key

7. Then server will encrypt the document again with decrypted secret key using Blowfish algorithm. Discard the decrypted secret key and original document

8. The Encrypted document, secret key and index will get stored on cloud server.



9. Now user wants to retrieve the document

10. User will give a search query; this query will get encrypted using user's private key and sent to server for search

11. Search query will decrypt by server and searched in index

12. Ranked results will get displayed to user

13. User will select a document d1, and then server will ask for secret key

14. If the secret key matches with key stored on server the user will get granted with the access to document and decrypted document will returned in response.

SYSTEM OVERVIEW:

The system architecture is concerned by creating a simple structural framework for a system. It defines the overall frame of the project which briefly describes the functioning of the structure and the purpose of the project phase is to plan a solution of the problem identified by the necessity file. The below Figure 1 shows the outline of the structure. We consider three parts in our system architecture: Data Owner, Data user and Cloud Server.

□ Data Owner is responsible for the creation of the database.

□ Data Users are the followers in a group who are able to use the files of the database.

□ Cloud Server deals information facilities to certified users. It is necessary that server be insensible to content of the database it keeps.

Data owner has amount of data records that he wishes to outsource on cloud server in encrypted form. Before outsourcing, data owner will first construct a secure searchable index from a set of diverse keywords removed from the file collection and store both the index and the encrypted file on the cloud server. We undertake the approval between the data owner and users is done. To search the file collection for a given keyword, certified user creates and submits a search request in a secret form-a trapdoor of the keyword to the cloud server. Upon getting the search request, the server is in charge to search the index and return the matching set of files to the user. We study the secure ranked keyword search problematic as follows: the search result must be returned giving to definite ranked relevance principles, to develop file retrieval correctness for users. Though, cloud server must study unknown or little about the important principles themselves as they reveal major sensitive information against keyword privacy. To decrease bandwidth, the user may send possible value k along with the trapdoor and cloud server only sends back the top-k most appropriate files to the user's concerned keyword.

PROBLEM STATEMENT :

Actually large number of on-demand data users and huge amount of data documents in the cloud, this difficulty is challenging. It is essential for the search facility to permit multi keyword search query and make available result comparison ranking to see the effective data retrieval requirement. To develop the search result accuracy as well as to enrich the user searching experience, it is also essential for such ranking system to support multiple keywords search, as single keyword search regularly yields extreme course results. The searchable encryption method supports to give encrypted data as documents and agrees a user to firmly search over single keyword and retrieve documents of concern.

CONCLUSION:

We describe and conclude the difficult of multi-keyword ranked search over encrypted cloud data, and create a variety of privacy requirements. Between numerous multi-keyword semantics, we select the effective similarity measure of "coordinate matching", i.e., as various matches as likely, to effectively capture the relevance of outsourced documents to the query communication .In our future work, we will search supporting other multi keyword semantics over encrypted data and checking the integrity of the rank order in the search result keywords. For agreement the challenge of supportive multi-keyword semantic without privacy breaks. Then we give two better MRSE outlines to persive many firm privacy requirements in two dissimilar threat models. Detailed analysis studying privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world data set show our future systems introduce low overhead on both computation and communication.

REFERENCES:

[1] Nilam Jamdare, Prof. K. V functions, "Privacy Preserving on multi-keyword search with Lucene Indexer over Encrypted Data in Cloud "

[2] Ms. Jabeen Akkalkot , Ms. S. Shanmug Priya "A SURVEY ON KEYWORDBASED SEARCH MECHANISM FOR DATA STORED IN CLOUD"

[3] Wenhai Sun, Bing Wang, Ning Cao , Ming Li "Privacy-preserving Multi-keyword Text Search in the CloudSupporting Similarity-based Ranking "

[4] Ning Cao , Cong Wang, Ming Li , Kui Ren , and Wenjing Lou "Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data"

[5] Mrs. Kavya B , Shrilakshmi P , Sushmitha N , Yamuna S . "Multi-Keyword Search Methodology for Cloud Data"

[6] Shiba Sampat Kale, Prof. Shivaji R Lahane "Privacy Preserving Multi-Keyword Ranked Search with Anonymous ID Assignment over Encrypted Cloud Data".