International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2494-9150 Vol-02, Issue 07, Oct 2016

Secure & Dynamic Multi-Keyword Ranked Search and Storing Encrypted Data on Cloud

[1]Alpesh Ahir, [2]Bhavika Gori, [3]Prof. Harsh Namdeo Bhor

[1,2]UG Student, [3]Asst. Professor, [1,2,3]K J Somaiya IE & IT, Sion, Mumbai, Maharashtra, India.

Abstract - As the cloud computing technology emerged during the last decade, outsourcing data to cloud service for storage becomes an attractive trend, which benefits in careful efforts on heavy data maintenance. since the outsourced distributed storage is not fully reliable, it raises security worries on the most proficient method to acknowledge information identical in cloud while accomplishing honesty inspecting. In this work, we examine the issue of uprightness evaluating and secure bit duplication on cloud information. In particular, going for accomplishing both information honesty and deduplication in cloud, we propose two secure frameworks, to be specific SecCloud and SecCloud+.

SecCloud presents a reviewing element with not keeping a MapReduce cloud, which helps the customers to produce information labels before transferring and in addition review the respectability of information have put away in cloud. Because of previous work the calculation by client in SecCloud is enormously decreased amid the document transferring and reviewing stages. SecCloud+ is outlined inspired by the way that customer dependably need to encode their information before transferring, and empowers respectability reviewing and secure deduplication on scrambled information.

Keywords: TF-IDF: Term Frequency - Inverse Document Frequency, greedy algorithm, g-divide-and-conquer algorithm, cost-model-based adaptive algorithm.

I. INTRODUCTION

Computing is being reworked to a model consisting of services that area unit commoditized and delivered in an exceedingly manner similar to utilities like water, electricity, gas, and telephony. In such a model, clients access administrations taking into account their prerequisites paying little heed to where the administrations are facilitated. A few processing ideal models have guaranteed to convey this utility processing vision. Cloud computing is that the most up-to-date emerging paradigm promises for showing the vision of 'computing utilities' into reality. A service providing computation resources is often named as Infrastructure as a Service (IaaS) and therefore the applications as computer code as a Service (SaaS).

An setting used for construction, deployment, and management of applications is called PaaS.

Distributed computing conveys framework, stage, and programming (application) as administrations, which are made accessible as membership situated administrations in a compensation as-you-go model to buyers. The value that CSPs (Cloud Service Providers) charge relies on upon the nature of administration (QoS) desires of CSCs (Cloud Service Consumer) Distributed computing encourages versatility and consistent adaptability of IT assets that are offered to end clients as an administration through the Internet. Distributed computing can help ventures enhance the creation and conveyance of IT arrangements by furnishing

them with access to administrations in a financially savvy and adaptable way.

Clouds can be classify into three categories, on the basis of their accessibility restrictions and the deployment model.

- Public Cloud,
- Private Cloud, and
- Hybrid Cloud.

An open Cloud is made accessible in a compensation as-you go way to the overall population clients independent of their starting point or alliance. A private Cloud's use is confined to individuals, workers, and trusted accomplices of the association.

A half and half Cloud empowers the utilization of private and open Cloud in a consistent way. Distributed computing applications traverse numerous areas, including business, innovation, government, wellbeing care, keen matrices, smart transportation systems, life sciences, fiasco administration, mechanization, information examination, furthermore, purchaser and informal organizations. Different models for the creation, sending, and conveyance of these applications as Cloud administrations have developed.

Cloud administration suppliers (CSPs) would guarantee to guarantee proprietors' information security utilizing purposes like virtualization what's more, firewalls. On the other hand, these instruments don't secure proprietors' information protection from the CSP itself, since the CSP holds full control of cloud equipment, programming, and proprietors' information. Encryptions on touchy information once sub-contracting can domain information protection alongside CSP. All things considered, information encryption sorts the conventional information usage administration in light of plaintext catchphrase look an exceptionally bewildering reprobate. A frivolous answer for this issue is to move all the encoded information and unscramble them close-by. Regardless, this technique is obviously impracticable since it will bring about a gigantic measure of correspondence overhead. Thus, rising a safe seek administration over scrambled cloud information is of abrogating noticeable quality. Secure inquiry over scrambled information has as of late pulled in light of a legitimate concern for some scientists.

II. LITERATURE SURVEY

In these section, presenting the different method to solve the problem related the cloud security:

Here they are established a set of difficult privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, they choose the efficient similarity measure of “coordinate matching”, i.e., to capture as many matches as possible relevance of data documents to the search query. Also further use “inner product similarity” to quantitatively evaluate such similarity measure. They first propose a basic idea for the MRSE based on secure inner product computation, and then give two improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. Thorough analysis identifying privacy and efficiency guarantees of proposed schemes is given. [1].

In another research proposed the searchable symmetric encryption scheme of an efficient similarity. To do so, they utilized locality sensitive hashing which is used for fast similarity search in high dimensional spaces for plain data. They proposed LSH based secure index and a search scheme to start fast similarity search in the context of encrypted data. In such a context, it is very critical not to loss the confidentiality of the sensitive data while providing functionality. They have provided a rigorous security definition and also proved the proposed scheme security under the provided definition to ensure the confidentiality. To clarify the details of the proposed scheme, we presented a real world application of it, namely the error identifying keyword search . Keyword search is enabled by this application which is tolerant to the typing errors both in the queries and the data sources. [2].

This research first exploits the popular similarity measure, i.e., vector space model with cosine measure, to effectively procure the accurate search result. They proposed two secure index schemes to meet various privacy requirements in the

two threat models. Eventually, the leakage of sensitive frequency information can be avoided.

To boost search efficiency, they propose a tree-based index structure for the whole document set. From the utilization of the prototype of our secure search system, identify three essential efficiency-related factors, by which the efficiency of the search algorithm upon our index tree can be significantly improved. In addition, whole search process made verifiable in case that users want to ensure the authenticity of the returned search results.[3]

This approach ensures that only the most related items are retrieved by the user, and also preventing unnecessary communication and computation burden on the user. System implements the whole system and demonstrates the effectiveness and efficiency of our solution through experiments using the publicly available Enron dataset. Our analysis depicted that the proposed scheme is proved to be secure, privacy-preserving, efficient and effective [4].

This research tackled the challenging problem of multi-keyword fuzzy search over the encrypted data. In that proposed and integrated several new designs to solve the problem of multiple keywords search and the fuzzy search simultaneously with high efficiency. Our approach of leveraging LSH functions in the Bloom filter to construct the file index is novel and provides an efficient solution to the secure fuzzy search of multiple keywords [5].

III. PROPOSED ARCHITECTURE

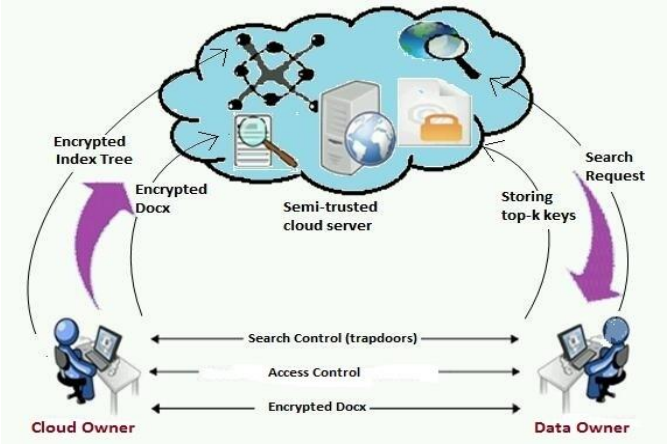

Fig.(1): Architecture Of Proposed System

Proposed system supports for both the accurate multi-keyword ranked search and flexible dynamic operation on document collection. MRSE is based on cloud but merging the concept of data mining. MRSE developed using AES encryption algorithm uses comparator interface for matching the strings.

New user can registered with One-Time-Password (OTP) which is very secure technique widespread used today.

In this section, our system provides the crucial steps of our proposed method. Search on encrypted cloud is performed through an encrypted searchable index that is generated by the data owner and outsourced to a cloud server. Given a query, server compares the query with the searchable index and returns the results without learning anything than the information that is allowed to be leaked due to efficiency concerns.

Index Generation

This proposed method utilizes the idea of bucketization which is a data partitioning technique widely used in literature. Here, each object is distributed into several buckets via min hash functions introduced in III-A and the bucket-id is used as an identifier for each object in that bucket. This method maps objects such that the number of buckets, in which two objects collide, increases as the similarity between those objects increases. In other words, while two identical objects collide in all of the buckets, number of common buckets decreases as dissimilarity between objects increases. The proposed secure index is generated by the data owner utilizing the following phases, named as feature extraction , bucket index construction and the bucket index encryption.

IV. ALGORITHMIC STRATEGY

AES Algorithm

The popular and widely adopted symmetric encryption algorithm likely to be encountered now a days is the Advanced Encryption Standard AES. It is found at least six times faster than triple DES. AES comprises three block ciphers, AES-128, AES-192 and AES-256. Every cipher encrypts and decrypts data in blocks of 128 bits using the cryptographic keys of 128-, 192- and 256- bits respectively.

(Rijndael's was designed to handle additional block sizes and key lengths, but the functionality were not adopted in AES.) Symmetric or secret-key ciphers use the same key for encryption and decryption , so both sender and the receiver should know and use the same secret key. All key lengths are deemed sufficient to protect important information up to the "Secret" level with "Top Secret" information requires either 192- or 256- bit key lengths. 10 rounds are there for 128-bit keys, 12 rounds for 192-bit keys, and 14 rounds for 256-bit keys – a round contains several processing steps that include substitution, transposition and mixing of the input plain text and then transform it into the final cipher text output. For MRSE implementation we use AES for the encryption method as well as decryption. Whenever user wants to upload their data on server it actually encrypt on users machine so that privacy is being preserved and data is safely stored. AES is working on background to performing encryption on entered data using encryption schemes and algorithm. AES is based on substitution-permutation network. It comprises of a series of linked three block ciphers. AES performs all its computations on

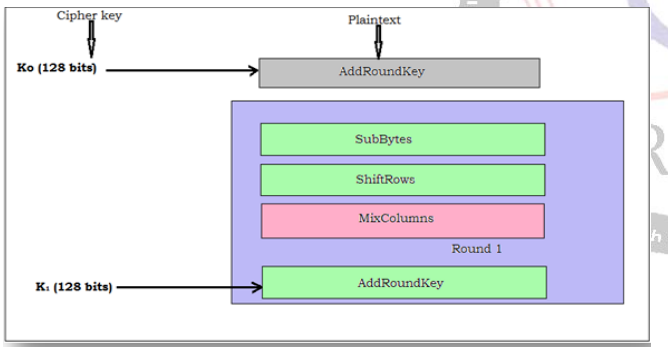

Fig.(2): AES Algorithm Working

Bytes rather than bits. AES treats the 128 bits of a plaintext block as 16 bytes. These 16 bytes are arranged in four columns and four rows for pressing as matrix. The number of rounds in AES is variable also it is depends on the length of key. In above figure 6.1.1 there is a description of actual round process.

Greedy DFS Algorithm

This algorithm construct a special structure of tree-based index and also propose a Greedy Depth-first Search algorithm to provide efficient multi-keyword ranked search . For obtaining high search efficiency, this algorithm construct a tree-based index structure and suggest a Greedy Depth-first Search algorithm based on this index tree. Due to the special structure of our tree-based index, the suggested search scheme can flexibly achieve sub-linear search time and deal with the deletion and insertion of documents.

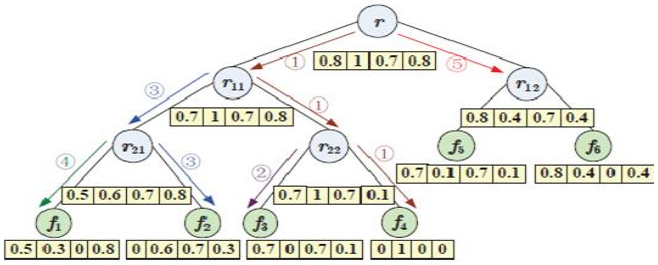

Fig.(3):- Greedy DSF

It is described depth-first search as estimating the promise of node n by a "heuristic evaluation function f(n) which, in general, may base on the description of n, the description of the goal, the information gathered by the search up to that point, and on any extra knowledge about the problem domain. "Some authors have used "depth-first search" to refer specifically to find with a heuristic that attempts to predict how close the end of a path is to a solution, so that paths are judged to be closer to a solution are extended first. This specific type of search is called Greedy Depth-first search or pure heuristic search.

Secure Search Scheme

To prevent different attacks in different threat models ,we construct two secure search schemes named as the basic dynamic multi-keyword ranked search (BDMRS) strategy in the Ciphertext model , and the enhanced dynamic multi-keyword ranked search (EDMRS) scheme in the known background model.

Searchable Encryption

Searchable encryption schemes allow the client to store the encrypted data to the cloud and execute keyword search over the cipher text domain. So far, under different threat models abundant works have been proposed to achieve various search functionality, like single keyword search, similarity search, multi-keyword Boolean search, ranked search, multi-keyword ranked search, etc. from them, multi-keyword ranked search achieves more and more attention for its practical

applicability. Recently, some dynamic strategies have been proposed to support inserting and deleting operations on document collection. These are significant works as it is possible that the data owners need to update their data on the cloud server. But few of the dynamic method support efficient multi-keyword ranked search.

V. CONCLUSION

In capable and secrecy Preserving Multi-Keyword Positioned Search over Encrypted Cloud Data, the precarious of secure multi-catchphrase scan for numerous information proprietors and various information clients in the distributed computing environment.

Particular from earlier works, these plans empower validated information clients to accomplish secure, catalyst, and adequate hunts more than a few information proprietors' information. To capably substantiate information clients and recognize assailants who take the mystery key what's more, execute illicit ventures, a novel element mystery key era convention and an imaginative information client verification convention is examined.

Particular from earlier works, these plans empower validated information clients to accomplish secure, catalyst, and adequate hunts more than a few information proprietors' information. To capably substantiate information clients and recognize assailants who take the mystery key what's more, execute illicit ventures, a novel element mystery key era convention and an imaginative information client verification convention is examined. To rank the query items and protect the security of pertinence scores between catchphrases furthermore, documents, we propose a novel Additive Order and Privacy Protecting Function family convention is examined.

In addition, it is demonstrated that the inclination is computationally compelling, notwithstanding for extensive information and catchphrase sets. The future work will consider the reprobate of secure fluffy watchword seek in a multi-proprietor worldview and to actualize the present plan on the suitable mists.

REFERENCES

[1] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, *Privacy-preserving multi-keyword ranked search over encrypted cloud data*, in IEEE INFOCOM, April 2011, pp.

[2] M. Kuzu, M. S. Islam, and M. Kantarcioglu, *Efficient similarity search over encrypted data, in Data Engineering* (ICDE), 2012 IEEE 28th International Conferenceon IEEE, 2012, pp. 11561167.

[3]W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, *Privacy preserving multi-keyword text search in the cloud supporting similarity-based ranking*, in Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security. ACM, 2013, pp. 7182.

[4] C. Orencik, M. Kantarcioglu, and E. Savas, *A practical and secure multi-keyword search method over encrypted cloud data*, in Cloud Computing (CLOUD), 2013 IEEE Sixth International Conference on. IEEE, 2013, pp. 390397.

[5] B. Wang, S. Yu, W. Lou, and Y. T. Hou, *Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud,* in IEEE INFOCOM, 2014.

