International Journal on Applications in Information and Communication Engineering
Volume 1: Issue 7: July 2015, pp 31-33. www.aetsjournal.com ISSN (Online) : 2394-6237

Secure Data Retrieval Using Efficient Similarity Measure Over Ranked Cloud Data

T.Niveathitha , S.Mohammed Jakkariya

Abstract— Cloud computing is rising as a promising pattern for knowledge outsourcing and high-quality knowledge services. However, issues of sensitive info on cloud doubtless cause privacy issues. Encryption protects knowledge security to some extent, however at the value of compromised potency. Searchable bi-radial encoding (SSE) permits retrieval of encrypted knowledge over cloud. For the primary time, the matter of privacy-preserving multi-keyword hierarchic search over encrypted knowledge in cloud computing (MRSE) has been resolved by establishing a collection of strict privacy needs for such a secure cloud knowledge utilization system. Among varied multi-keyword linguistics, the economical similarity live of "coordinate matching," i.e., as several matches as potential, to capture the connexion of knowledge documents to the search question is chosen any use "inner product similarity" has been accustomed quantitatively value such similarity live. Here basic plan for the MRSE supported secure real computation, so offer considerably improved MRSE schemes to attain varied demanding privacy needs in 2 completely different threat models.to boost search expertise of the information search service, these 2 schemes is extended. Thorough analysis investigation privacy and potency guarantees of planned schemes is given. Experiments on the real-world knowledge set any show planned schemes so introduce low overhead on computation and communication.

Keywords: Cloud computing, Searchable bi-radial encoding, MRSE,

I. INTRODUCTION

Cloud computing could be a technology that uses the net and central remote servers to keep up knowledge and applications. Cloud computing permits customers and businesses to use applications while not installation and access their personal files at any pc with web access. This technology permits for rather more economical computing by consolidative storage, memory, and process and information measure. Cloud computing could be a comprehensive resolution that delivers IT as a service. The flexibleness of cloud computing could be an operate of the allocation of resources on demand. Before cloud computing, websites and server-based applications were dead on a selected system. Cloud computing is softened into 3 segments application, storage and property.

T.Niveathitha, PG Student , Dept of Computer Science & Engineering, Shanmuganathan Engineering College, Pudukkottai. (Email: niveathitha.ktn@gmail.com)

S.Mohammed Jakkariya, Asst.Professor, Dept of Computer Science & Engineering, Shanmuganathan Engineering College, Pudukkottai. (Email: jakkariya@gmail.com)

II. RELATED WORK

A. SEARCHABLE ENCRYPTION SINGLE KEYWORD

Single keyword searchable encoding schemes sometimes build associate encrypted searchable index specified its content is hidden to the server unless it's given acceptable trapdoors generated via secret keys. The searchable encoding construction, wherever anyone with public key will write to the information keep on server however solely licensed users with non-public key will search. Public key solutions area unit sometimes terribly computationally overpriced but. more over, the keyword privacy couldn't be protected within the public key setting since server may inscribe any keyword with public key so use the received trapdoor to guage this ciphertext.

B. BOOLEAN KEYWORD SEARCHABLE ENCRYPTION

To enrich search functionalities, conjunctive keyword search over encrypted knowledge are planned. These schemes incur giant overhead caused by their elementary primitives, like computation price by additive map. As a a lot of general search approach, predicate encoding schemes area unit recently planned to support each conjunctive and dividing search. Conjunctive keyword search returns "all-or-nothing," which implies it solely returns those documents during which all the keywords given by the search question appear; dividing keyword search returns uniform results, which implies it return seaRch document that contains a set of the particular keywords, even only 1 keyword of interest.

III. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

- Today's mail servers like IMAP servers, file servers and different knowledge storage servers usually should be totally trusted—they have access to the information, and thus should be trusty to not reveal it while not authorization—which introduces undesirable security and privacy risks in applications.
- Previous work shows a way to build encrypted file systems and secure mail servers, however usually one should sacrifice practicality to confirm security. The elemental drawback is that moving the computation to the information storage appears terribly tough once the information is encrypted, and plenty of computation issues over encrypted knowledge antecedently had nosensible solutions.
- The trivial resolution of downloading all the information and decrypting regionally is clearly impractical, as a result of the massive quantity of information measure price in cloud scale systems.

• Considering the doubtless {large number sizable quantity of on-demand knowledge users and big amount of outsourced knowledge documents within the cloud, this drawback is especially difficult because it is very tough to satisfy conjointly the wants of performance, system usability and measurability to satisfy the effective knowledge retrieval want, the big quantity of documents demand the cloud server to perform result connexion ranking, rather than returning uniform results.

• Such hierarchic search system allows knowledge users to seek out the foremost relevant information quickly, instead of burdensomely sorting through each match within the content assortment.

• Ranked search may elegantly eliminate superfluous network traffic by causing back solely the foremost relevant knowledge, that is extremely fascinating within the "pay-as-you use" Cloud paradigm.

• One of the foremost fashionable ways in which to try to to therefore is thru keyword-based retrieval. Keyword-based retrieval could be a typical knowledge service and wide applied in plaintext eventualities, during which users retrieve relevant files during a file set supported keywords.

• It seems to be a tough task in cipher-text state of affairs as a result of restricted operations on encrypted knowledge.

• A series of searchable encoding (SSE) schemes are planned to change search on cipher-text. ancient point schemes change users to firmly retrieve the cipher-text, however these schemes support solely mathematician keyword search, i.e., whether or not a keyword exists during a file or not, while not considering the distinction of connexion with the queried keyword of those files within the result.

• Top-k multi keyword over encrypted cloud knowledge. These schemes, however, suffer from problems—Boolean illustration and the way to strike a balance between security and potency.

LIMITATIONS

• Boolean illustration and the way to strike a balance between security and potency.

• Security and privacy risks.

• Only provide Single keyword Search.

• Top-k multi-keyword has been used however solely provide mathematician search.

• Data lekage happens

3. 2 PROPOSED SYTSTEM

• Solve the matter of multi-keyword graded search over encrypted cloud information (MRSE)where as protective strict systemwise privacy within the cloud computing paradigm.

• The economical similarity live of "coordinate matching," i.e., as several matches as doable, to capture the connexion of information documents to the search question.

• Use "inner product similarity" i.e., the quantity of question keywords showing in an exceedingly document, to quantitatively assess such similarity live of that document to the search question.

• During the index construction, every document is related to a binary vector as a subindex wherever every bit represents whether or not corresponding keyword is contained within the document.

• The search question is additionally delineated as a binary vector wherever every bit suggests that whether or not corresponding keyword seems during this search request, that the similarity may well be precisely measured by the scalar product of the question vector with the info vector.

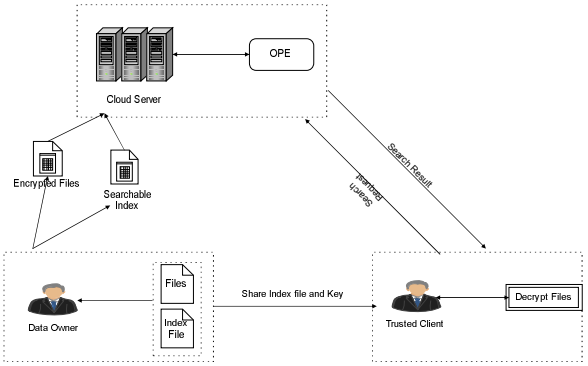
• A basic plan for the MRSE exploitation secure scalar product computation, that is customized from a secure k-nearest neighbour (kNN) technique then offer 2 considerably improved MRSE schemes in an exceedingly gradual manner to attain varied rigorous privacy needs in 2 threat models with enlarged attack capabilities.

ADVANTAGES

• The new theme guarantees high information privacy.
• Less communication overhead.
• Provide significant security for storage.

IV. SYSTEM ARCHITECTURE

The data owner includes a assortment of information documents F to be outsourced to the cloud server within the encrypted kind C. To change the looking capability over C for effective information utilization, the info owner, before outsourcing, can 1st build associate encrypted searchable index I from F, then source each the index I and therefore the encrypted document assortment C to the cloud server. to look the document assortment for t given keywords, a certified user acquires a corresponding trapdoor T through search management mechanisms, as an example, broadcast secret writing. Upon receiving T from an information user, the cloud server is accountable to look the index I and come back the corresponding set of encrypted documents. to boost the document retrieval accuracy, the search result ought to be graded by the cloud server in step with some ranking criteria (e.g., coordinate matching).



V. ALGORITHM - MRSE FRAMEWORK

The MRSE system consists of four algorithms as follows:

- Setup
- BuildIndex
- Trapdoor

- Query

Setup (1 [l]): Taking a security parameter " *l*" as input, the data owner outputs a symmetric key as SK.

BuildIndex (F, SK): Based on the data set F, the data owner builds a searchable index I which is encrypted by the symmetric key SK and then outsourced to the cloud server. After the index construction, the document collection can be independently encrypted and outsourced.

Trapdoor ($w\tilde{}$): With t keywords of interest in $w\tilde{}$ as input, this algorithm generates a corresponding trapdoor $T_{w\sim}$.

Query ($T_{w\sim}$, k, I): When the cloud server receives a query request as ($T_{w\sim}$, k), it performs the ranked search on the index I with the help of trapdoor $T_{w\sim}$, and finally returns $F_{w\sim}$, the ranked id list of top-k documents sorted by their similarity with $w\tilde{}$.

VI. PERFORMANCE ANALYSIS PRECISION AND PRIVACY

Therefore, similarity scores of documents will be not exactly accurate. In other words, when the cloud server returns top-k documents based on similarity scores of data vectors to query vector, some of real top-k relevant documents for the query may be excluded. This is because either their original similarity scores are decreased or the similarity scores of some documents out of the real top-k are increased, both of which are due to the impact of dummy keywords inserted into data vectors. To evaluate the purity of the k documents retrieved by user, we define a measure as precision Pk ¼ k0=k where k0 is number of real top-k documents that are returned by the cloud server.

EFFICIENCY

i) Index Construction

To build a searchable subindex Ii for each document Fi in the data set F, the first step is to map the keyword set extracted from the document Fi to a data vector Di, followed by encrypting every data vector. The time cost of mapping or encrypting depends directly on the dimensionality of data vector which is determined by the size of the dictionary, i.e., the number of indexed keywords. And the time cost of building the whole index is also related to the number of subindex which is equal to the number of documents in the data set.

ii) TRAPDOOR GENERATION

The time to generate a trapdoor is greatly affected by the number of keywords in the dictionary. Like index construction, every trapdoor generation incurs two multiplications of a matrix and a split query vector, where the dimensionality of matrix or query vector is different.

Size of Subindex/Trapdoor

Size of dictionary	4000	6000	8000	10000	12000
MRSE_I (KB)	31.3	46.9	62.5	78.1	93.8
MRSE_II (KB)	32.5	48.1	63.8	79.4	95.0

VII. CONCLUSION

Among varied multi-keyword linguistics, the economical similarity live of "coordinate matching," i.e., as several matches as doable, to effectively capture the connexion of outsourced documents to the question keywords, and use "inner product similarity" to quantitatively assess such similarity live. Thorough analysis in work privacy and potency guarantees of planned schemes is given, and experiments on the real-world dataset show the planned schemes introduce low overhead on each computation and communication. The integrity of the ordering within the search result's being cheched exploitation TPA(Third Party Auditor) assumptive the cloud server is un-trusted. Index Hash Table is built to avoid information modification and colliding attacks.

REFERENCES

[1] D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2004.

[2] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Processing Private Queries over untrusted clod data through privacy homomorphism ," Proc. IEEE INFOCOM, pp. 829-837, Apr, 2011.

[3] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized Private Keyword Search over Encrypted Data in Cloud Computing," Proc. 31st Int'l Conf. Distributed Computing Systems (ICDCS '10), pp. 383- 392, June 2011.

[4] J. Katz, A. Sahai, and B. Waters, "Predicate Encryption Supporting Disjunctions, Polynomial Equations, and Inner Products," Proc. 27th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2008.

[5] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy Keyword Search Over Encrypted Data in Cloud Computing," Proc. IEEE INFOCOM, Mar. 2010.

[6] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.