ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Energetic Data Maintenance with Encryption and Privacy Preserving Nature over Cloud Environment

[1]Jyothsna.S, [2]Dr Jitendranath Mungara, [3]Gangadhar C.Immadi, [4]Vandana CP,

Department of Information Science, New Horizon College of Engineering, India

Professor & Head, Department of Information Science, New Horizon College of Engineering, India

Sr.Assistant Professor, Department of Information Science, New Horizon College of Engineering, India

Sr.Assistant Professor, Department of Information Science, New Horizon College of Engineering, India

ABSTRACT: Now-a-days IT industry is blasting. In this way, the prerequisite for assets is likewise on the expansion. Industry requires additionally handling power and capacity ability to meet their objective. Here, Cloud Processing comes in the photo, it gives IT industry the genuinely necessary assets on a substantial scale easily and makes their errand simple. Associations can without much of a stretch outsource their enormous measure of information to distributed storage. In any case, the protection of information is a major concern. The information security can be accomplished by encryption procedures; however it expands the trouble of safely seeking information on the cloud in light of the fact that seeking in scrambled information is itself a testing undertaking. As of late many plans have been proposed yet, these plans don't consider the semantic of the question. We proposed a novel strategy by consolidating LSI and hierarchical group to get the semantic connection between the outcomes and to lessen the pursuit space separately. Further, to check the inquiry result validness, we utilize MAC tree alongside a cryptographic signature. Through security and execution investigation we demonstrate that our technique is superior to anything past encoded searchable plans.

KEYWORDS: Searchable Encryption, Secure Data Maintenance, Dynamic Multi-Keyword Searching Scheme, Cloud Computing.

I. INTRODUCTION

Recently, Searchable Encryption [SE] frameworks have been delivered for secure outsourced data look. Some further investigates focus on chase capability multi-catchphrase look for and secure component reviving .But they simply support adjust watchword look. To overhaul the interest versatility and convenience, some examination has been done on feathery watchword look for. These game plans support versatility of minor linguistic missteps and design abnormalities, for instance, search for "million" by means of neglectfully stating "million", or "data mining" by expressing "data mining". These arrangements generally take the structure of terms into thought and use change detachment to survey the similarity.

They didn't consider the terms semantically related to question catchphrase, in this way many related records are blocked. Moreover, these cushy systems send back each imperative record only upon closeness/nonattendance of the watchword, and result-situating is still out of considering. In this system, from another perspective, we propose an equivalent interest course of action in perspective of semantic question augmentation while supporting similarity situating. Semantic advancement based tantamount chase fortifies the system accommodation by giving back the decisively organized reports and the records including the terms semantically related to the request catchphrase.

In the proposed plot, a looking at archive metadata is worked for each record. By then the mixed metadata set and record social occasion are exchanged to the cloud server. With the metadata set, the cloud server collects the reworked record and constructs Semantic Relationship Library [SRL] for the watchwords set. The co-occasion of terms is used to evaluate the semantic association between terms in SRL. Subsequent to getting a question request, the cloud server thus finds the terms which are semantically related to the request catchphrase as showed by the estimation of semantic association between terms in SRL.

By then both the catchphrase and the semantically stretched out words are used to recoup records. Finally, the planned archives are given back all together as showed by the total significance score. At the same time, to ensure security and last result situating, we fittingly change a grave primitive demand shielding encryption to guarantee the importance score. Distinct security examination shows that the plan viably understanding the target of semantic interest, while ensuring the security. Wide trial evaluation shows the adequacy and reasonable of the arrangement.

II. EXISTING SYSTEM

Researchers have proposed various ciphertext look for contrives by intertwining the cryptography methods. Likewise, the association between records is shrouded in the above techniques. The association between documents addresses the properties of the reports and from now on keeping up the relationship is key to totally express a record. For example, the relationship can be used to express its class. In case a document is self-sufficient of some different reports except for those records that are related to recreations, then it is basic for us to validate this record has a place with the order of the diversions. Because of the outwardly hindered encryption, this fundamental property has been canvassed in the ordinary procedures. Thusly, proposing a procedure which can keep up and utilize this relationship to speed the chase stage is appealing. Sun et al. use Merkle hash tree and cryptographic check to make an unquestionable MDB-tree. Regardless, their work can't be particularly used as a piece of our plan which is arranged for security ensuring various catchphrase look for. In this way, a genuine framework that can be used to affirm the question things inside gigantic data circumstance is crucial to both the CSPs and end customers.

Disadvantages

Starting late, a couple request arranges have been proposed, for instance, single watchword and various catchphrase in perspective of Boolean model and Vector space show separately to ensure data security with efficient investigate mixed data. Each one of these arrangements could look for data in perspective of watchwords however the issue with catchphrase chase is that they have substitute significance in different setting or various words have relative ramifications, which gives screwed up results when addressed.

In this way, it is basic to get the right significance of a word and for what it is used as a piece of that record. Existing procedures have been shown with provable security, however their methods require huge operations and have high time diserse quality. In this manner, past techniques are not fitting for the immense data circumstance where data volume is gigantic and applications require online data taking care of.

- Melody et al. system has a high looking for cost in view of the analyzing of the whole data amassing word by word.
- Because of the nonattendance of rank instrument, customers need to set aside a long chance to pick what they require when colossal reports contain the request watchword. Along these lines, the demand protecting frameworks are utilized to comprehend the rank segment,
- Sun et al. give another outline which fulfills better chase viability. In any case, at the period of rundown building process, the significance between files is neglected.

III. PROPOSED METHODOLOGY

A vector space model is utilized and each record is spoken to by a vector, which implies each archive can be viewed as a point in a high dimensional space. Because of the connection between various archives, every one of the reports can be partitioned into a few classifications. In our framework demonstrate, there are three elements: the information proprietor, the information client, and the cloud specialist organization. The information proprietor first records report and structure the various leveled bunch. At that point encodes both report set and the various leveled bunch index and outsources it to the distributed storage. The information proprietor likewise validates the information client to guarantee that lone verified clients are permitted to get to the information. The cloud server gives the required storage room and gives the item to information proprietor and information client in light of the hunt question. Both Data proprietor and Data client are trusted, and CSP is semi-trusted.

Advantages

The hunt time can be generally decreased by choosing the coveted class and relinquishing the unessential classifications. Contrasting and every one of the reports in the dataset, the quantity of archives which client goes for is little. Because of the modest number of the coveted records, a particular classification can be additionally isolated into a few sub-classes.

A virtual root is built to speak to every one of the information and classifications. The virtual root is meant by the hash consequence of the link of the considerable number of classifications situated in the main level. The virtual root will be marked with the goal that it is obvious. To confirm the query item, client just needs to check the virtual root, rather than confirming each archive.

Protection Preserving through MRSE: This venture go for comprehending the security issues in the cloud. The cloud's extraordinary adaptability and monetary funds are rousing banks to outsource their nearby complex information administration framework into the cloud. To secure information protection and battle spontaneous gets to in the cloud and past, touchy information, charge reports, money related exchanges, et cetera, may must be scrambled by information proprietors before outsourcing to the business open cloud. Nonetheless, this obsoletes the customary information use benefit in light of plaintext watchword seek.

The minor arrangement of downloading every one of the information and decoding locally is plainly unrealistic, because of the gigantic measure of transmission capacity cost in cloud scale frameworks beside disposing of the neighborhood stockpiling administration, putting away information into the cloud fills no need unless they can be effortlessly sought and used. Along these lines, investigating protection saving and powerful inquiry benefit over encoded cloud information is of vital significance. From one viewpoint, to meet the powerful information recovery require, the vast measure of archives request the cloud server perform result importance positioning, rather than returning undifferentiated outcomes.

Such positioned seek framework empowers information clients to locate the most significant data rapidly, instead of difficult dealing with each coordinating the substance accumulation.

IV. LITERATURE SURVEY

In the year of 2012 the authors "Jia Zhi-Peng ; Zhang Ya-Ling ; Wang Shang-Ping ; Sun Qin-Dong", described into their paper titled "Evaluable Secure Ranked Keyword Search Scheme over Encrypted Cloud Data" such as a sensitive information is centralized into the cloud, how to effectively search over encrypted data is becoming a research bottleneck of Cloud Storage Security. Based on the original solution of secure ranked keyword search over encrypted cloud data, we exploit the concept of evaluation for relevance of index. We introduce the Index Evaluation Service to build a secure ranked keyword search over encrypted cloud data. With the user's subjective evaluation for relevance of index, our solution achieves the goal of efficient utilization of the search result and user's feedback.

In the year of 2010 the authors "Dawn Xiaoding Song ; D. Wagner ; A. Perrig", described into their paper titled "Practical techniques for searches on encrypted data" such as: It is desirable to store data on data storage servers such as mail servers and file servers in encrypted form to reduce security and privacy risks. But this usually implies that one has to sacrifice functionality for security.

For example, if a client wishes to retrieve only documents containing certain words, it was not previously known how to let the data storage server perform the search and answer the query, without loss of data confidentiality. We describe our cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems. Our techniques have a number of crucial advantages.

They are provably secure: they provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext when only given the ciphertext; they provide query isolation for searches, meaning that the untrusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching, so that the untrusted server cannot search for an arbitrary word without the user's authorization;

they also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server.

The algorithms presented are simple, fast (for a document of length n, the encryption and search algorithms only need O(n) stream cipher and block cipher operations), and introduce almost no space and communication overhead, and hence are practical to use today.

In the year of 2010 the authors "Ning Cao ; Cong Wang ; Ming Li ; Kui Ren ; Wenjing Lou" described into their paper titled "3. Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data" such as With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search.

Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE).

We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we choose the efficient similarity measure of "coordinate matching," i.e., as many matches as possible, to capture the relevance of data documents to the search query. We further use "inner product similarity" to quantitatively evaluate such similarity measure.

We first propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. To improve search experience of the data search service, we further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given. Experiments on the real-world data set further show proposed schemes indeed introduce low overhead on computation and communication.

Information Retrieval Process

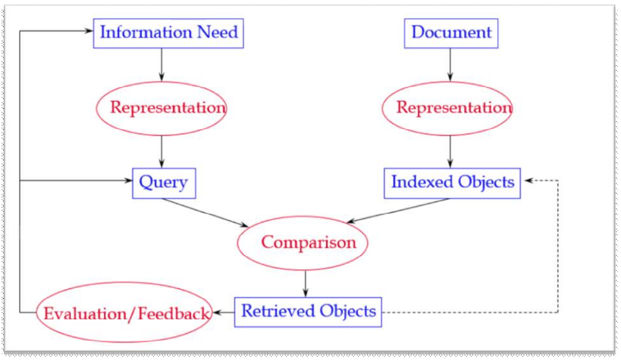

Fig.1 Information Retrieval

Boolean Model: Boolean model, we have categorized solutions into two forms:

(a) Single Keyword Encrypted Text Search Scheme: Song et al. proposed the first Searchable Encryption (SE) Scheme to search in the encrypted data. They used a symmetric key for encryption and decryption, but it has some security issues. To improve the security, Boneh et al. proposed a new Public key encryption technique to search in encrypted data. Main drawbacks of above schemes were that they do not consider the relevance between the query and retrieved documents. These schemes were purely based on Boolean model and work with single keyword only. It makes it inefficient for use.

(b) Multi-Keyword Encrypted Text Search Scheme: To solve the problems of single keyword text search, Cao et al ,proposed a multi-keyword encrypted text search scheme. They used coordinate matching and produced ranked result by the number of matched keywords. Many other similar schemes were proposed such as Ballard et alHu et al. but the main drawback of all these proposed schemes was, that they are based on Boolean model, so no partially matched documents are retrieved. These schemes are not able to make a balance between security and efficiency. Also, the user must have a good domain knowledge before forming a query.

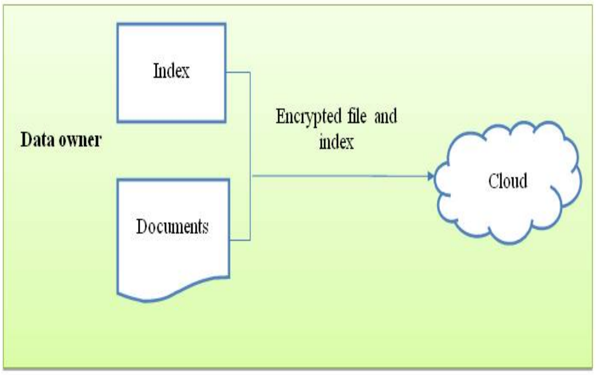

Fig.2 Data Owner Process Flow

Vector Space Model: we have categories these solutions into two forms:
(a) Single Keyword Encrypted Text Search Scheme: Pang et al. proposed the first scheme for encrypted text search based on vector space model and used cosine similarity to measure the similarity between query and documents and put a threshold value for similarity measure. Only those documents are retrieved, which cross the threshold value. Thus, only most relevant documents are retrieved, this scheme increases the accuracy of the system. The main drawback of this model was that it was single keyword encrypted text search scheme and the user cannot search a query with multi-keywords.

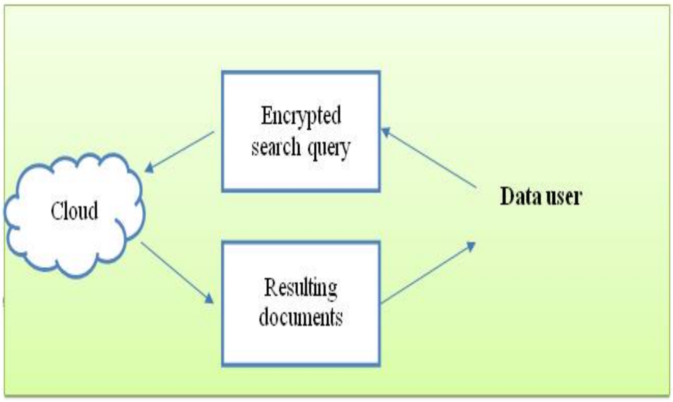

Fig.3 Data User Process Flow

(b) Multi-Keyword Encrypted Text Search Scheme: Sun et al. proposed a new scheme based on vector space model which supports multi-keyword encrypted text search. It was similar to previous schemes, but it is capable of searching multi-keyword encrypted text based queries which will ensure increased accuracy and efficiency of the system. J. Yu et al. also proposed a multi-keyword search over encrypted text search scheme, and also many other schemes were proposed under multi-keyword search scheme, such as Cao et alSun et abut the main drawback with vector space model is that it does not capture the semantic of a document, so there is a possibility that search result may have documents based on the query keywords but not related to the concept of the query. The user may form a query based on some idea in mind, but not know what keywords to include in it. Those keywords may have some different meanings in different contexts.

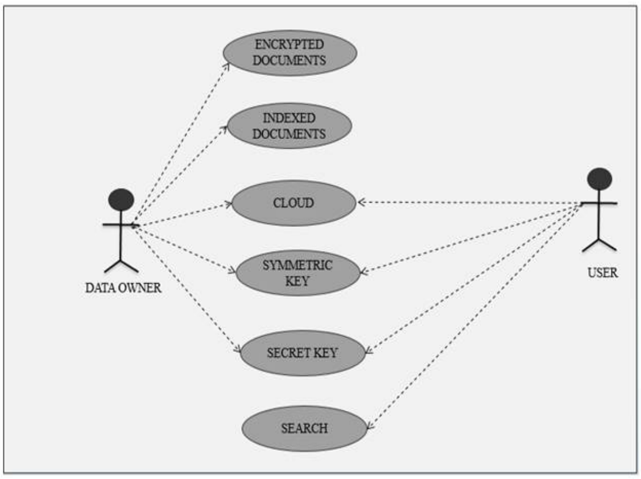

Fig.4 System UseCase Diagram

V. SEARCH RESULT VERIFICATION SCHEMES

There is a possibility of software or hardware failure, which may cause wrong results or there might be some internal employees, who modified or changed the stored encrypted documents or part of documents. So, to verify the obtained results integrity, Sun et al. [proposed a scheme in which they used a secured tree based structure. The Cloud server returns minimum sub-tree and user run the same search algorithm on that minimum sub-tree to authenticate the search query. The basic drawbacks of these schemes are that they required more computation time and storage space to process large size tree structure, which increases the extra overhead.

VI. CONCLUSION

In this framework, we tended to the issue of information protection and analyzed existing arrangements. At that point proposed An Efficient Privacy-Preserving Multi-watchword Ranked Search over Encrypted Data in Cloud Computing, which utilizes LSI and various leveled grouping to recover more related archives and increment the efficiency of the inquiry over encoded information in the cloud condition. We utilized the semantic connection amongst archives and the question to accomplish the efficient look, which is all around positioned in view of the inquiry watchwords with protecting information security. We additionally utilized the MAC technique to confirm the trustworthiness of the information. Through security and execution examination, we have demonstrated that EPPMRSE is secure and efficient than existing plans. In future, we will take a shot at element informational indexes.

REFERENCES

1.N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data,” Proc. IEEE INFOCOM, pp. 829-837, Apr, 2011.

2. L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, “A Break in the Clouds: Towards a Cloud Definition,”. ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50-55, 2009.

3. N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, “LT Codes-Based Secure and Reliable Cloud Storage Service,” Proc. IEEE INFOCOM, pp. 693-701, 2012

4. S. Kamara and K. Lauter, “Cryptographic Cloud Storage,” Proc. 14th Int’l Conf. Financial Cryptography and Data Security, Jan. 2010.

5. A. Singhal, “Modern Information Retrieval: A Brief Overview,” IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35-43, Mar. 2001.

6. I.H. Witten, A. Moffat, and T.C. Bell, Managing Gigabytes: Compressing and Indexing Documents and Images. Morgan Kaufmann Publishing, May 1999.
7. D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.
8. E.-J. Goh, "Secure Indexes," Cryptology ePrint Archive, http://eprint.iacr.org/2003/216. 2003.
9. Y.-C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data," Proc. Third Int'l Conf. Applied Cryptography and Network Security, 2005.
10.R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions" Proc. 13th-2006.