An Enhanced Multi - Keyword Ranked Search using Co-Ordinate Matching for Privacy Preserving over Encrypted Cloud Data

S.Amsaveni[1] G.T.Prabavathi[2]
[1]Research Scholar [2]Associate Professor
[1,2]Department of Computer Science
[1,2]Gobi Arts & Science College, Gobichettipalayam

Abstract— Due to the tremendous growth of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for flexibility and economic savings. Search service for the encrypted cloud data is essential. Allowing multiple keywords in the search request and returning documents in the order of their relevance to the keywords is an interesting area of research. Preserving the privacy of the sensitive data is an important issue for cloud computing. This paper discusses about the challenging problem of privacy preserving using Multi-keyword Ranked Search (MRSE) over encrypted cloud data. The proposed work focus on searchable encryption using multi-keyword search and Co-ordinate matching keyword search, and sort the search results. The proposed system is tested and the result shows a considerable amount of improvement is searching time. The retrieval of documents is also significantly higher than MRSE.

Key words: Co-Ordinate Matching for Privacy Preserving, Encrypted Cloud Data, Multi - Keyword Ranked Search

I. INTRODUCTION

Cloud refers to a Network or Internet which provides services over public and private networks. Cloud computing refers to manipulating, configuring, and accessing the hardware and software resources remotely (5). It offers online data storage, infrastructure, and has numerous advantages to access applications as utilities, over the Internet and can manipulate and configure the applications online at any time, In Cloud computing, information flow easily and securely to and from the user and these valuable resources must be saved securely and be effectively backed up and protected from disasters.

In cloud computing, a data owner checks the files in the system and index the outsourced data to cloud server and request send to the server and it searches the required file mentioned in the ranking ordered. When the request send to the cloud server, search request and file may easily encrypt the data in file. Considering a cloud data hosting service involves the three different entities, the data owner, the data user along with the ID, and the cloud server. The data owner first registers on cloud using anonymity algorithm for cloud computing service. Before saving user registration information to database present on cloud anonymous algorithm process the data and then anonymous data is saved to registration database. The data owner has a collection of data documents F to be outsourced to the cloud server in the encrypted form C. To enable searching capability over C for effective data utilization, the data owner has to build an encrypted searchable index I from F before outsourcing , and then outsource both the index I and the encrypted document collection C to the cloud server. The cloud computing deals with efficient algorithms for assigning identifiers to the users on the cloud in such a way that the IDs are anonymous using a distributed computation with no central authority (9). In order to improve the document retrieval accuracy, the search result should be ranked by the cloud server according to some ranking criteria such as Co-ordinate matching and assigning anonymous ID to the user on cloud in order to make the data on cloud more secure.

Privacy Preserving is an important issue for cloud computing both in terms of legal complains and user trust and needs to be considered at every phase of design. Privacy preserving is used to preserve the security of fields (18). If a database has to be shared among several users and some data contained in the database should be prevented by using access control methods in order to guarantee that only authorized people are allowed to have access that sensible information, then the need of privacy and preserving the privacy emerges. Several users can share their databases and some data are protected by using access control methods. It allows only authorized people to have access sensible information.

The aim of privacy preserving algorithm is the extraction of relevant knowledge from large amount of data, while protecting at the same time sensitive information. Several Data mining techniques incorporating privacy protection mechanism have been developed that allow one to hide sensitive item sets or patterns, before the data mining process is executed. In information retrieval community, there exist state-of-the-art techniques to achieve good result ranking using multi-keyword queries on plain text. The success of privacy preserving data mining algorithm is measured in terms of its performance and data level of uncertainty. Next chapter deals with ranked keyword searching and the necessity of multi-keyword ranked searching.

II. RANKED SEARCH

A. Keyword Based Search:

This type of search is essential in cloud data and typically done in such a way that the users can utilize cloud data to query a collection (19). Through keyword searching, users can selectively retrieve files of interest and has been widely applied in plaintext search scenarios. To eliminate unnecessarily network traffic by not sending back the irrelevant data, ranked keyword search is used. Applying keyword based search in the encrypted cloud data is a challenging task due to security and privacy obstacles. This technique is highly desirable in the "pay-as-you-use" cloud paradigm. For privacy protection, such ranking operation should not leak any keyword related information. Multi-keyword ranked search algorithms are readily available in information retrieval process and it is essential for cloud

server data while protecting the cloud data as well as to enhance the search privacy.

Ranked keyword searching help the user to get the accurate result based on the multi-keyword concepts. The users can enter the multiple words query, the server splits that query into a single word after search that word file in the database. Finally, display the matched word list from the database and the user gets the file from that list. The search query is also described as a binary vector where each bit means whether corresponding keyword appears in this search request, so the similarity could be exactly measured by inner product of query vector with data vector.

The traditional information retrieval has already provided multi-keyword ranked search for the data user. The similar scenario should be applied to cloud server. Cloud storage should allow various keywords in a query and the present the data documents in the relevant order. The main aim of this work is to find the solution of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm.

B. MRSE

The MRSE proposed by Ning et al. (6) preserves privacy in cloud to meet the privacy requirement and prevent the cloud server from learning additional information from data mining. The MRSE requirements are: (i) Index Confidentiality: The keyword values of keywords are stored in the index. Thus, the index stored in the cloud server needs to be encrypted (2) Trapdoor Unlink ability: The cloud server could do some statistical analysis over the search result. Meanwhile, the same query should generate different trapdoors when searched twice. The cloud server should not be able to deduce relationship between trapdoors (3) Keyword Privacy: The cloud server could not discern the keyword in query, index by analyzing the statistical information like term frequency.

III. RELATED WORK

Shiba Sampat Kale et al. (9) suggested that it is very important to take privacy into account while designing cloud services, if these involve the collection, processing or sharing of personal data. In this algorithm, the authors build a searchable inverted index that stores a list of mapping from keywords to the corresponding set of files which contain this keyword. When data users input a keyword, a trapdoor is generated for this keyword and then submitted to the cloud server. The above stated algorithm combined inverted index with order-preserving symmetric encryption (OPSE). In terms of ranked search, the order of retrieved files is determined by numerical relevance scores, which can be calculated by TF×IDF. The relevance score is encrypted by OPSE to ensure security.

Song et al. (11) developed MDB-tree based scheme which supports ranked multi-keyword search. This methods employ a spell check mechanism. An efficient fuzzy keyword search over encrypted data in cloud computing was proposed by Jin et al. (4) which used edit distance to build fuzzy keyword sets. Bloom filters are constructed for every keyword. Then, it constructs the index tree for all files where each leaf node has a hash value of a keyword. The edit distance is used to quantify keywords similarity and construct storage efficient fuzzy keyword sets. Specially, the wildcard-based fuzzy set construction approach is designed to save storage overhead. In the searching phase, if the edit distance between retrieval keywords and ones from the fuzzy sets is less than a predetermined set value, it is considered similar and returns the corresponding files.

In the algorithm proposed by Li et al. (5) a new scheme named Latent Semantic Analysis (LSA)-based multi-keyword ranked search which supports multi-keyword latent semantic ranked search is used. By using LSA, the proposed scheme could return not only the exact matching files, but also the files including the terms latent semantically associated to the query keyword. For example, when the user inputs the keyword “automobile” to search files, the proposed method returns not only the files containing “automobile”, but also the files including the term “car”. In the LSA based algorithm, the authors took a large matrix of term-document association data and construct a semantic space wherein terms and documents are closely associated are placed near to one another. To meet the challenge of supporting such multi-keyword semantic without privacy breaches, the authors proposed the idea: the multi-keyword ranked search (MRSE) using “Latent Semantic Analysis.

To meet the challenge of supporting such multi-keyword semantic without privacy breaches, the author (5) proposed a basic SMS scheme using secure inner product computation, which is adapted from a secure k-nearest neighbor (KNN) technique, and then improve it step by step to achieve various privacy requirements in two levels of threat models: 1) Showing the problem of Secured Multi-keyword search over encrypted cloud data and 2) Propose two schemes following the principle of co-ordinate matching similarity.

Ning Cao et al (7) established a set of strict privacy requirements for solving Privacy Preserving MRSE in Cloud Computing. They proposed a basic idea for the MRSE based on secure inner product computation. In their proposed technique “Coordinate Matching” is used as similarity measure to capture relevant data for a search query. The “Inner Product Similarity” scheme is used to quantitatively evaluate the similarity measure. The algorithm is practical, flexible and has low overhead on both computation and communication. In the MRSE scheme proposed by the authors, a new random number t is assigned to the extended dimension of each query vector to increase the difficulty for the cloud server to learn the relationships among received trapdoors. In addition a dummy keyword is inserted in each data vector and a random value is assigned to it. The authors also proposed a more advanced MRSE scheme to achieve various privacy requirements in two different threat models.

The author solves the challenging problem of privacy preserving MRSE over encrypted cloud data based on secure inner product computation and efficient similarity measure of coordinate matching, i.e., as many matches as possible in order to capture the relevance of data documents to the search query. The authors proposed significantly improved MRSE scheme to achieve various string and privacy requirements in two different threat models. An algorithm for anonymous sharing of private data among N parties is developed. This technique is used iteratively to assign these nodes ID numbers ranging from 1 to N. This

assignment is anonymous in that the identities received are unknown to the members of the group.

An efficient multi-keyword search is proposed by Yanzhi et al (19). The author proposed a light weight search that supports efficient multi-keyword ranked search in cloud computing system. First a basic scheme using polynomial function is used to hide the encrypted keyword and search patterns for efficient multi-keyword ranked search. To enhance the search privacy, the authors proposed a privacy preserving scheme which utilizes the secure inner product method for protecting the privacy of the searched multi-keywords. In this work, the authors encrypted the keywords and then constructed a polynomial function to hide them in search trapdoor generation. The detailed scheme to achieve the ranked multi-keyword search over encrypted data are performed using the functions setup, build index and trapdoor. The authors used ranked multi-keyword search scheme to hide the keywords in the search query using polynomial functions which provides guaranteed privacy over two threat models. Through the proposed scheme, the authors conducted extensive experiments based on the real-world dataset. The experimental results demonstrated that the scheme proposed by Yanzhi et al (19) enables the encrypted multi-keyword ranked search service is highly efficient in cloud computing.

IV. PROPOSED SYSTEM

A ranked scheme with KNN and Triple Data Encryption Algorithm (TDES) is proposed in this work. From number of multi-keyword semantics highly efficient rule of co-ordinate matching is selected through as many matches as possible to identify the similarity between search query and data. In the proposed system a basic Secured multi-keyword ranked search scheme using secure inner product computation is proposed and then it is improved to meet different privacy requirements. The Ranked result provides top k retrieval results. The following steps are required while searching takes place in encrypted data using multi-ranking keyword search: (i) Multi-dimensional query are converted to its secure multi-keyword search using (SMS) Co-ordinate matching (ii)Attributes are defined in a hierarchical way. i.e., attribute hierarchy (iii) Indexes and capabilities are generated by MRSE and average precision is calculated respectively.

The keyword index is encrypted using pseudo random bits using heuristic pseudo random function. On the setup phase, the user chooses a random secret input to encrypt the file. Then the user submits index and file content to server. On the retrieval phase, when the users wants to retrieve file from the server, user retrieves the index file and then compute keyword with the secret input. The K-nearest neighbor search identifies the top k nearest neighbors to the query matched on user cloud to retrieve the result based on co-ordinate matching technique.

A. Triple DES Algorithm

TDEA provides a relatively simple method of increasing the input size of DES to protect against attacks, without the need to design a completely new block cipher algorithm. The algorithm is given below:

1) Triple DES uses a "key bundle" that comprises three DES keys, K1, K2 and K3, each of 56 bits (excluding parity bits).
2) Encrypt: Convert cipher text = EK3 (DK2 (EK1 (plaintext))). i.e., DES encrypts with K1, DES decrypt with K2, then DES encrypt with K3,etc.
3) Decryption in the reverse: Plaintext = DK1 (EK2 (DK3 (cipher text))), i.e., decrypt with K3, encrypt with K2, and then decrypt with K1.
4) Each triple encryption encrypts one block of 64 bits of data. In each case the middle operation is the reverse of the first and last. This improves the strength of the algorithm when using keying option 2, and provides backward compatibility with DES with keying option 3.

The designed algorithm is implemented, tested and the results are analyzed. The implementation issues and results are discussed in next chapter.

V. RESULTS AND DISCUSSIONS

In this work, three sets of experimental queries are used to investigate dictionary of 2050 words is created and placed on cloud which is used for searching. The comparison of MRSE and co-ordinate matching accuracy is shown in Figure 1.

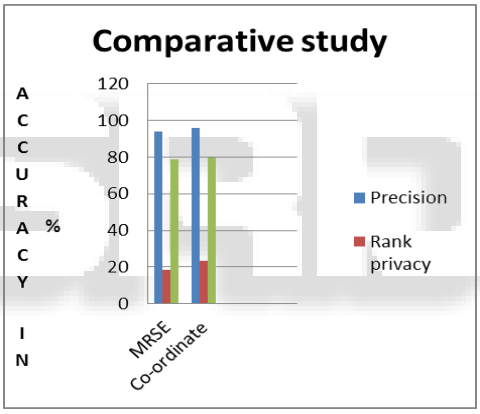

Fig. 1: Accuracy

The below Figure 2 is a Comparison graph of MRSE and Co-ordinate matching (KNN and TRIPLE DES). The graph plots the number of documents with respect to system search result returned and time required to return the documents. The system requires less time which is less than around 30 documents with most specific result of number of document equal to one which is less than the documents returned by MRSE system which are required is around 20 documents.

Figure 3 shows a Comparison graph of MRSE and Co-ordinate matching (KNN and TRIPLE DES). The graph plots the number of Documents that the respective system's search result returned and Time required to return the documents in respective System. The graph shows how the propose algorithm requires less time which is less than around 9 minutes with most specific result of number of document equal to one which is less than the documents returned by MRSE system which are three and time required is around 6 minutes. The algorithm achieves effective ranking result using KNN and Triple DES technique. The results of the proposed system shows a significant improvement in terms of document retrieval as well as time taken to retrieve the documents. This system currently

works on single cloud. In future, work is will extended up to remote computing and provides better security in multi-user systems.

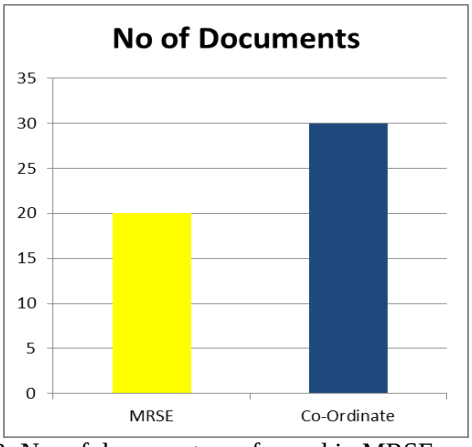

Fig. 2: No. of documents performed in MRSE and Co-ordinate matching

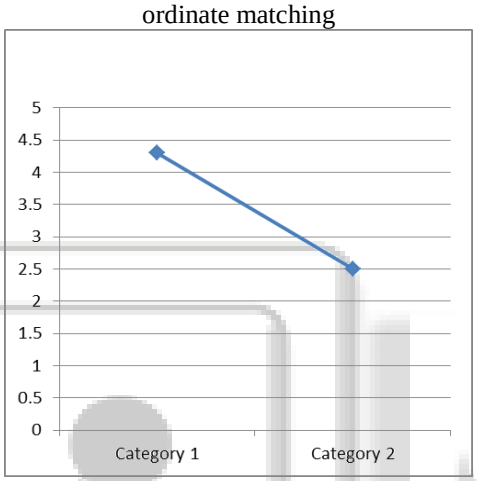

Fig. 3: Time of query in MRSE and Co-ordinate matching.

VI. CONCLUSION

Searching a document from the cloud is as interesting area of research and various techniques have been proposed to preserve privacy in cloud computing. The research community has developed many privacy preserving algorithms to extract relevant knowledge from large amount of data, while protecting the time sensitive information. As large number of users attempt to retrieve files from cloud, keyword base search is the commonly used technique to search over the encrypted cloud data. Multi-keyword ranked searching helps to the user to get the accurate result based on multi keywords.

The main objective of this research work to provide an improved multi ranked search. Co-ordinate matching is a measure that uses the number of query keywords appearing the document to quantify the relevance of the document the query. Hence, in this research Co-ordinate matching is taken for searching the cloud data. For evaluating the performance of the proposed system, a cloud setup was done and then Triple DES approach with KNN is applied to search the documents from the encrypted cloud data. In the future, the encrypted data of semantic keyword search using LSA can be designed. Further, privacy preserving keyword searches on remote encrypted data can be performed.

REFERENCES

[1] Archana, S and K. Elangovan, "Survey of Classification Techniques in Data Mining", International Journal of computer science and mobile application, Vol.2,Issue.2, pp.65-71,2014

[2] Cong Wang, Ning Cao, Jin Li, Kui Ren, and Wenjing Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," in Proceedings of ICDCS, Vol.2, pp.253-262,2010.

[3] Igor Nai Fovino and et al., "State of the Art in Privacy Preserving Data Mining," JRC Scientific and Technical Reports, Jan.2008

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

[5] Li Chen, Xingming Sun, Zhihua Xia , Qi Liu, "An Efficient and Privacy Preserving Semantic Multi-Keyword Ranked Search over Encrypted Cloud Data" International Journal of Security and Its Applications, Vol.8, No.2, pp.323-332, 2014.

[6] Ning Cao, Z. Yang, Cong Wang, Kui Ren and Wenijing Lou, "Privacy preserving Query over Encrypted Graph-Structured Data in Cloud Computing, in Proceedings of Distributed Computing Systems CDCS), pp.393-402, 2011.

[7] Ning Cao, Cong Wang, Ming(Fred) Li, Kui Ren, and Wenijing Lou, "Privacy preserving Multi-Keyword Ranked Search over Encrypted Cloud Data", in Proceedings of IEEE INFOCOM, Vol.25,No.1, pp. 829– 837, 2014.

[8] Rakesh Agrawal and Ramakrishnan Srikant, "Privacy Preserving data mining", proceedings of the ACM SIGMOD Conference on Management of Data, pp. 439–450,May 2000.

[9] Shiba Sampat Kale,Prof.Shivaji R Lahane, "Privacy preserving Multi-Keyword Ranked Search with Anonymous ID Assignment over Encrypted Cloud Data" International Journal of Computer Science and Information Technologies, Vol.5 (6), pp.7093-7096, 2014.

[10] Shuceng Yu, Cong Wang, Kui Ren and Wenjing Lou," Achieving Secure, Scalable and Fine-Grained Data Access Control in Cloud computing," in Proceedings. IEEE INFOCOM, pp. 535-542, 2010

[11] Song. D, D. Wagner, and A. Perrig. "Practical techniques for searching on encrypted Data". In IEEE Symposium on Research in Security and Privacy, page 44-55. IEEE Computer Society, 2000.

[12] Sun.W, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Verifiable Privacy Preserving Multi-keyword Text Search in the Cloud Supporting Similarity-based Ranking", IEEE Transactions on Parallel and Distributed Systems, 2014.

[13] Vanishree R ,Mr.G.S Suresh, "Multi-Keyword Ranked Search Over Encrypted Cloud Data", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)Vol.4 Issue5, May 2015.

[14] Vimmi Makkar, Sandeep Dalal, "Ranked Keyword Search in Cloud Computing: An Innovative Approach",

International journal of computational engineering research\\ Vol.03\\ Issue 6, June 2013.

[15] Wenhai Sun, Bing Wang, Ning Cao, Ming Li, Wenjing Lou, Y. Thomas Hou, Hui Li, "Privacy-Preserving Multi-Keyword Text search in the cloud supporting similarity-based ranking", in Proceedings of ACMSIGSAC, pp.253-262,2013.

[16] Wang.B, Y. Hou, M. Li, H. Wang, and H. Li, "Maple: Scalable multidimensional range search over encrypted cloud data with tree- based index", in Proceedings of ACM ASIACCS, 2014.

[17] Wong.W.K, D. W.-l. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases", in Proceedings of the ACM SIGMOD International Conference on Management of data, 2009.

[18] Yan-Cheng, Chang and Michael Mitzenmacher. " Privacy preserving keyword Searches on remote encrypted data". IProceedings of ACNS'05, pp.442–455, 2005.

[19] Yanzhi Ren, Yingying Chen, Jie Yang, and Bin Xie, "Privacy preserving Ranked Multi- Keyword Search Leveraging Polynomial Function in Cloud Computing "Privacy preserving Ranked Multi-Keyword Search Leveraging Polynomial Function in Cloud Computing", Globeco., IEEE, pp.594 - 600, 2014.

