Multi-Keyword Privacy Security Protected Search Through Encrypted Data On Cloud Storage

Anwar Basha H[1], S. Sasi Kumar[2], D Dhanasekaran[3]

[1]Research Scholar, CSE Department
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India
Email : anwar.mtech@gmail.com
[2]Professor, CSE department, Saveetha Engineering College,
Thandalam, Chennai, India
sasikumar@saveetha.ac.in
[3]Professor, CSE Department,
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India
nddsekar@gmail.com

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

Abstract:: Through the advancement of cloud storage, data providers are allowed to outsource their complicated data processing processes with considerable versatility and economic benefits since local locations to the private, public cloud. However, confidential data must be secured before outsourcing in order for data to be safe, which obsolescent common usage of data relying on plaintext search. Therefore, it is incredibly necessary to allow a secure cloud data search service Given the vast number of web users and cloud records, in order to be applicable to such keywords, several keywords should be acknowledged in the search request and return records. The searchable encryption-based research specializes in search of a particular keyword or in boolean keywords, seldom processing the outcomes of pursuit. For the very first time, the problem of protecting the privacy of multi - keyword search for encrypted data in cloud storage is described as well as solved in this article. For such a safe cloud data usage scheme, we series a series of strict privacy criteria. We pick from the different multi-keyword semantics the effective "coordinate matching" similarity test, i.e. as various matches as likely, in order to catch the importance of the data papers to the search. We can use "inner product similitude" to test this resemblance attribute quantitatively. Initial, on the basis of stable internal commodity estimation and then, in two separate Vulnerability Models, we suggest a simple concept for the MRSE to reach multiple rigorous data protection standards. We expand these two frameworks to help further search terminology to improve our user experience for the data search service. Examination of data security and assurance of performance of suggested systems shall be carried out in-depth. Experiments in the real-world data set indicate that plans for numerical and connectivity structures effectively add small overheads.

Keywords: Privacy, cloud computing, key, ranked search

1 Introduction:

Cloud infrastructure is the long imagined idea enabling cloud users to access their information in the cloud securely in order to experience high - quality software along with facilities from a collective pool of configurable computing power, on-demand as well as on-demand. All people and organizations are inspired to move their local dynamic data processing infrastructure into the cloud with considerable versatility and economic savings. Critical information, for instance, e - mails, individual health reports and picture archives, fiscal documentation and financial transactions, etc., that needs to be protected by data proprietors in order to defend the protection of information and counter illegal accesses in the cloud along with beyond. Nevertheless, this obsolescent the conventional search for keywords focused on the plaintext. Thanks to the immense latency costs of cloud-scale applications, the simple approach to store and decode all data locally is obviously unrealistic. In fact, the collection of cloud data does not have a function because local resource control is removed, and it can be accessed and used easily. It is also highly necessary to pursue privacy protection and powerful search service over encrypted cloud records. In the cloud, this challenge is especially difficult because of a theoretically vast number of consumers with data on request and an immense amount with external data documentation and consistency, device accessibility and scalability criteria are often incredibly hard to satisfy.

At the one side, the large volume of documents allows the cloud provider to rate performance significance instead of returning undifferentiated outcomes in order to satisfy the successful data recovery requirements. This confidential search method helps data users to easily locate the most important material, rather than to filter every match burdensomely in the selection of contents. Rankings will elegantly remove unnötige network traffic by returning just the most appropriate results, which in the cloud "pay-as-you" model is highly attractive. Nonetheless, for privacy security, no keyword-based details can be released. At the other side, it is often important for these ranking systems to promote several searching keywords in order to increase search efficiency as well as optimize user quality, as searching for one keyword sometimes results in way too gross results. Information consumer will choose to have a series of keywords, as a standard procedure suggested in the existing site search engines (e.g. Google, Bing), rather than three, as the representation of their seeking interest for the most important information. And any keyword in the seek question will help you further refine the search result.

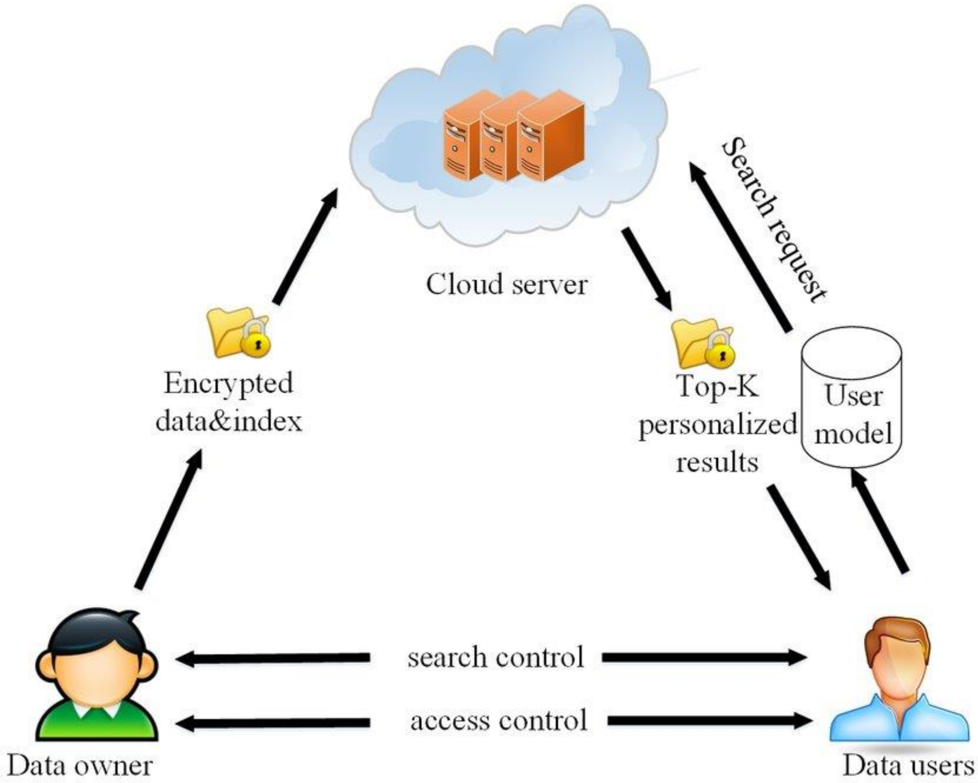

Figure. 1 Architecture of the search through encrypted data on cloud

We propose a simple framework intended for MRSE utilizing safe in-line information computation to address the task of promoting such multi-keyword semantics without breaches of privacy and then include two dramatically enhanced MRSE schemes, one with step – by - step approaches to satisfy different strict privacy demands in two threats models through an increase.

The list of our efforts is:

1. We investigate the multi - keyword search question for encrypted cloud information for the earliest time and develop a collection of strict privacy standards for a safe cloud service customer.

2. In two separate models of thematic challenges, we give two MRSE systems focused on the similarity test of the "coordinate matching."
3. We are looking at some more changes in our categorized search system in order to accommodate further search chemicals and complex results.
4. Thorough analyzes are carried out on the privacy and performance assurances of the initiatives and real-world data set tests to demonstrate that the proposed solutions effectively impose small overheads on computing and connectivity.

This article provides two different frameworks to help more search semantics relative to the previous edition of this paper. This edition also discusses the help of system architecture for data/index dynamics. In addition, the study and review of 2 different techniques would boost the experimental research. We provide further information on the protected internal component and the privacy aspect in addition to these enhancements.

2 Related work:

The resulting rating was [1] created to get better the accuracy of the search results along with to enhance the user experience in searching. This proposed scheme cuts the time of retrieval, i.e. the time to locate the documents needed, by 90% and minimizes the total upgrade[2] database burden as latest files have to be updated (database generation time) as opposed to the offered effective indexing schemes in writing for a comparable dataset. This has not traditionally been done by maximizing the quest period along with index generation time. We often use internal commodity seamlessness to determine the seamlessness attribute quantitatively. First, to achieve a multi-keyword protected scan, we suggest a fundamental[3] concept for MRSE focused on safe internal information estimation and then have two substantially better mechanisms for MRSE to meet specific privacy criteria. We evaluated the attack in-depth in this paper and also presented systematic proof that the attack is indeed weak.[4] A modern, effective, reliable, multi-keyword search framework that enables the web users to access the web safely from the untrusted public cloud is therefore necessary. It is vital. We evaluate protection to explain the consistency and privacy of the planned schemes. [5]Comprehensive real-world dataset tests authenticate our research along with proof that our proposed approach supports synonym-based scanning very effective and accurate.

A tree dependent directory through a preservative arrange as well as privacy-preservingng set of functions is designed for each data owner to achieve an efficient search.[6] This cloud service will then effectively combine such indexes by searching the corresponding files with the algorithm of "Depthfirst Quest." Finally, the strict safety investigation demonstrates that our scheme is secure, and the reliability and quality of our success measurement were shown. We are introducing a novel class of additives and secrecy protection instructions. We suggest a new hybrid secret key production protocol along with a new information user verification procedure to deter attackers since eavesdropping secret keys and claiming to be legitimate computer users conducting searches. [7]In fact, PRMSM facilitates secure deletion of device consumers. Comprehensive real-world data set tests affirm the PRMSM's efficacy and performance.

The review of protection and results shows that MRSE-HC, along with EMRSE-HC, preserve the secrecy of multi - keyword protected search [8] schemes for hybrid clouds. The weight of the-IDF and the weight of choice help ensure that the findings remain accurate to the value of the consumer. Consequently, detailed analyzes of protection and[9] efficiency on real-world data set tests have shown that PPSE does indeed comply with our architecture goals. These two principles protect the security of data and guarantee the safety of cloud consumers. In addition, it demonstrates that proposed[10] solutions incorporate quick recovery, stability and lower costs in public cloud storage and communication.

The observations indicate that the search period for the new approach decreases linearly with a significant rise in records in the data and the search period for the standard method rises exponentially.[11] In fact, the new approach has an advantage over the conventional system in the secrecy and validity of the records obtained. Searchable script encryption results in a multi-keyword question consistency over authenticated cloud information as well as returns the best applicable top-k performance. [12] wide-ranging experimental data sets demonstrate that the suggested solution will minimize index capacity dramatically and improve recovery performance in real-time. A thorough study of the safety and reliability of the proposed model has been carried out, and tests found that both measurement and interaction have obtained a low overhead of the proposed model.

3 Proposed method:

Throughout corresponding literature, such as searchable encryption, reflective protection promises that the user does not even know search data. We discuss and set up a series of strict privacy criteria, especially for the MRSE system, with this general secrecy definition.

In order to conduct a multi-keyword test, we suggest that the "internal product resemblance" is used to determine the successful resemblance metric in the quantitative way "coordinate match": first we suggest a fundamental concept for MRSE through the use of a stable internal product measurement adapted by a safe KNN process. More search terminology and dynamic application are also provided.

MRSE_I: Privacy-Preserving method in recognized Ciphertext representation

Our MRSE concept is not strong enough for the adapted, safe inner commodity measurement scheme. More notably, the only alleged factor concerned is the balance factor r in the trapdoor series, which does not offer sufficiently ambiguity overall as needed by the necessity for trapdoor unlinkability and the privacy necessity for keywords. We at this time sell our MRSE I scheme to have a more sophisticated interface for MRSE.

MRSE_I Scheme

In our additional sophisticated architecture, we retain this expanding procedure but add a latest casual t to the complete dimension in-question vector instead of merely deleting the enlarged dimension from the question vector, as we expect to do at first glance. This modern randomness will find it impossible to understand the relationship between the trapdoors for the cloud service. In addition, randomness should be carefully controlled, as defined in the privacy criteria of the keyword in the search results, to impede the frequency of the document and minimize the possibilities for keyword reidentifying.

MRSE_II Scheme

The above privacy leaks are triggered by the random variable' I' of the Data Vector Di. More dumb keywords as an alternative of just one should be introduced into each Di data vector to delete such a fixed property of every single Di text.

MRSE_I_TF

The keyword inclusion in the text or the database is seen as 1 in the matrix or the database variable in the "coordinate matching" rating theory. There are even several considerations that may influence the efficiency of the quest. For instance, in mainly documents in the data gathering, where one keyword is present in the report, the value of that keyword is smaller than that of other keywords contained in fewer documents. Similarly, if a document includes several query keywords, then the user can prefer the query keyword to the other document in one location only.

MRSE_II_TF

Here, though a few entries in Di have modified the scale analysis assault introduced in Section 4.3 from binary value 1 to normalized phrase occurrence, still partly functions in the defined history model. The document rating will then be transmitted to the cloud repository and then used in the established history model to define this keyword.

4 Result and discussion:

Within this portion, a detailed theoretical evaluation of the suggested methodology on an existing data collection is demonstrated: the Enron Email Data Collection. The competence of four anticipated MRSE schemes, together with the balance among search accuracy along with privacy, have been evaluated for our technology. As seen in Section 4, dumb keywords are embedded in every information vector, along with some are chosen in every question. Comparison ratings in records would also not be identical. This implies that any actual top-k related document for the database may be omitted as the cloud repository returns top-k documents dependent on the resemblance scores of data vectors for the application vector. That is when the original resemblance values in certain records are either diminished, or similarity ratings are raised, all attributed to the results in dumb keywords introduced into data vectors. We describe the metric as precision Pk 1/2 k0= k to determine the pureness of the documents received by the consumer while k0 is the number of actual top-k documents returned by the cloud server.

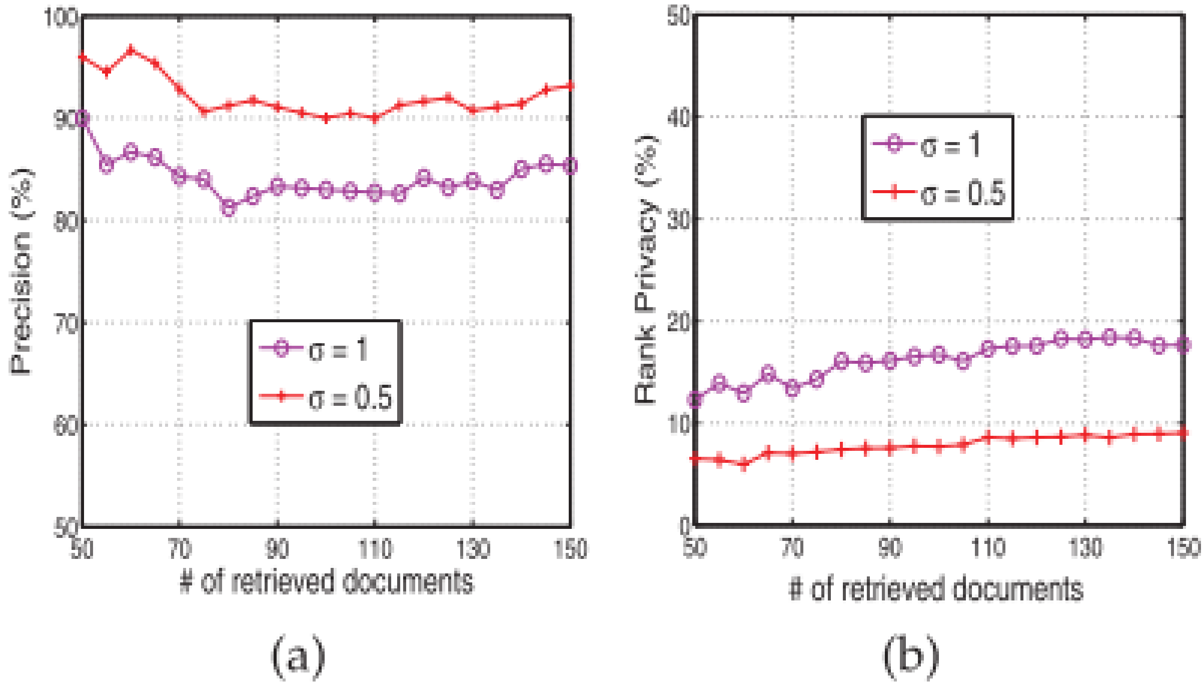

Figure.2 There is a discrepancy among (a) precision also (b) rank privacy with a particular selection of the standard deviation for the random variables.

Figure. 2a suggests facilitating the MRSE quality is obviously influenced by the standard differs from the chance changeable.' Standard deviation is supposed to be smaller in terms of efficiency in order to attain strong consistency showing the decent integrity of the records obtained.

Efficiency

Index Construction

The initial move is to map the keyword collection extracted from the Fi document to a data vector di in order to construct the searchable subindex Ii for each Fi document in the data set format. The mapping and/or encryption costs depend explicitly on the dimension of the data vector, calculated by dictionary scale, i.e. the number of keywords that have been indexed. And the time-consuming creation of the whole index often relies on the number of subindexes relative to the number of documents on the data collection.

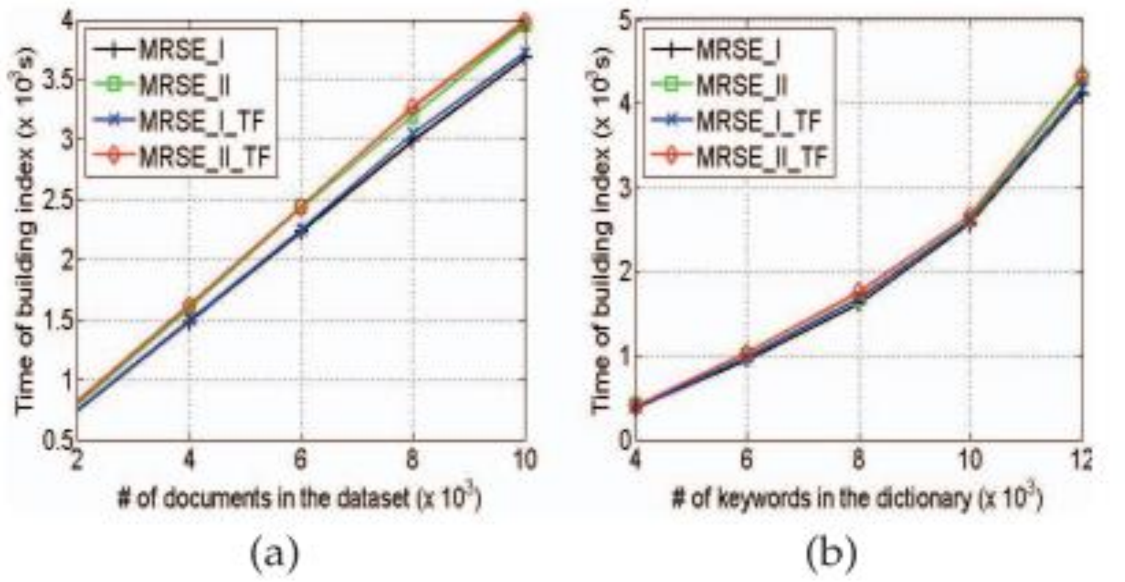

Figure. 3 The time cost of building index. (a) For the different size of the data set with the same dictionary, (b) For the same data set with different size of the dictionary

. Figure. 3b implies that the amount of keywords indexed in the dictionary dictates the time expense of a subindex creation. Introduce further computation during the index creation as the term frequency information has to be obtained on all records on each keyword, and then standardized measurements are done. Yet, as shown in the statistics, even a further calculation is negligible in the TF IDF weighting law, since even further calculation comes from separating and multiplication of the matrix. While the period of production index for the data owner is not a marginal cost, it is a one-time process until data outsourcing. The sizes of the subindex are fully compatible with the dimensionality of the data function when the number of keywords in the dictionary defines. Owing to marginal variations in data vector size, subindex measurements are very similar in the two MRSE schemes.

Trapdoor Generation

Figure. 4a reveals that the amount of keywords in the vocabulary significantly affects the time to construct a trapdoor. Like index structures, each iteration of the trapdoor engenders two matrix multiplications as well as a split query vector in which in two schemes the matrix or query vector dimensionality becomes specific and decreases as the dictionary size grows.

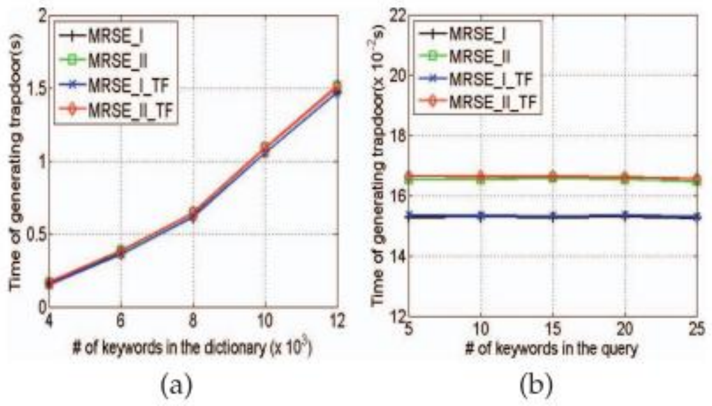

Figure 4. Trapdoor processing period bill (a) Keywords in different dictionary sizes with the same question. (b) For different question keyword numbers in the same dictionary

Figure. 4b reveals that in the MRSE II method, the cost to manufacture trapdoors is around 10 per cent higher with sophistication than in the MRSE I method. The discrepancy between MRSE I TF along with MRSE II TF is comparable when the additional logarithm approximation is very low for the complete trapdoor series. The difference in costs to build trapdoors, as with subindex production, is mainly due to the vector and matrix measurements of the two MRSE systems. More importantly, the number of query keywords has little impact in the overhead generation of trapdoors, an advantage over associated multi-keyword scanning.

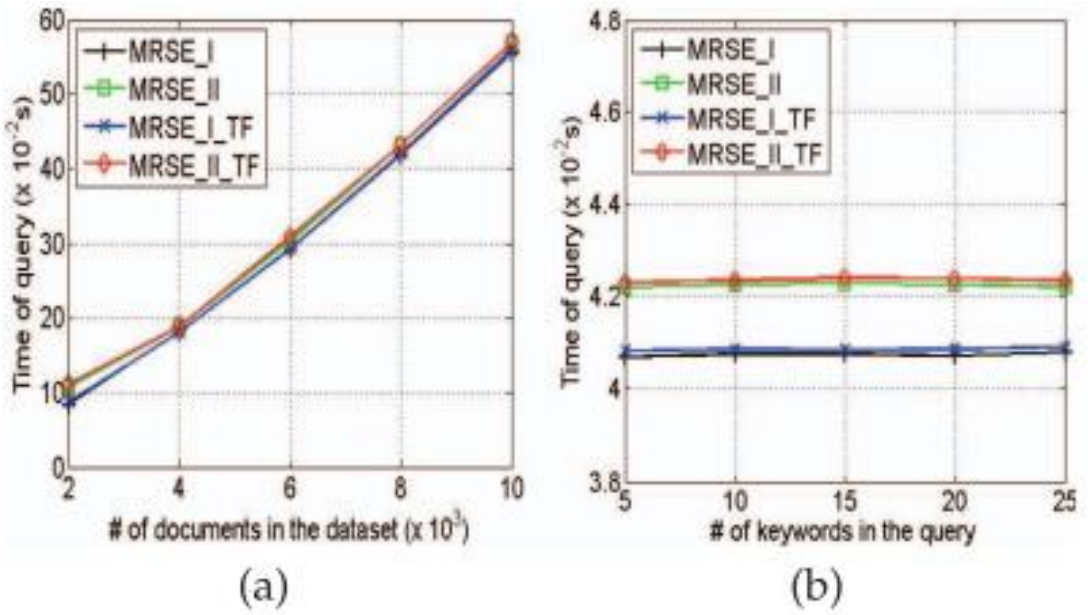

Figure. 5. Question time fee. (a) Keywords in different data collection sizes for the same question. (b) for different question keyword numbers inside the same data collection.

Query

Server execution for a cloud-server involves the calculation and ranking of correlations across certain data sets. Similarity values in data selection MRSE I and MRSE I TF are calculated and strengthened in TF ratings MRSE II and MRSE II. The answer period depends on the data scale, although the quantity of keywords in question has an extraordinarily limited impact on the results, the cost of trapdoor output above. Figure 5 shows the response time. The two schemes in the established MRSE I and MRSE I TF variant of the ciphertext model are very close in size as they have the same measurements which are the main deciding factor for the expense of computations.

5 Conclusion:

This article is the first time that the multi-keyword search issue is identified and discussed over encrypted cloud data, and a set of privacy standards are created. We have chosen the efficient similarity calculation of coordinate communication across specific multi-keyword semantics, i.e. to catch as many matches as possible the significance of outsourced documents to the keywords of querying and use "inner object similitude" to quantify this similarity quantitatively. We propose a simple concept of MRSE utilizing stable inner component approximation in order to address the task of promoting multi-keyword textual without privacy breaches. For two separate vulnerability models, we instead introduce two modified MRSE systems to satisfy a set of strong privacy criteria. We also discuss several potential changes to our search system, including support for further search semanticity, i.e. TF IDF, and dynamic data operations. Comprehensive analyzes are performed on the secrecy and functionality of schemes introduced. Real-world data sets evaluations demonstrate that our schemes provide small overhead for calculation and connectivity. We must investigate the credibility of the rating in our future research, given that the cloud provider is not trusted.

Reference:

1. Rane, Deepali D., and V. R. Ghorpade. "Multi-user multi-keyword privacy preserving ranked based search over encrypted cloud data." In 2015 International Conference on Pervasive Computing (ICPC), pp. 1-4. IEEE, 2015.
2. Anwar Basha H, Arunnehru J, Sathya R, Meenakshi. "Multi Keyword Ranked based Search for Secured Cloud Data using Vector Space Model." In 2020 Test Engineering and Management (TEM), pp. 14228 – 14234. March – April 2020 (Volume 83), 2020.
3. Roshini Rajendran, Vani V Prakash. "An Efficient Ranked Multi-Keyword Search for Multiple Data Owners Over Encrypted Cloud Data: Survey". In 2018 International Research Journal of Engineering and Technology (IJERT), pp. 1382 – 1386, Volume: 05, Issue: 09, 2018.
4. Krishna, C. Rama, and Sneha A. Mittal. "Privacy preserving synonym based fuzzy multi-keyword ranked search over encrypted cloud data." In 2016 International Conference on Computing, Communication and Automation (ICCCA), pp. 1187-1194. IEEE, 2016.
5. Jivane, Anjali Baburao. "Time efficient privacy-preserving multi-keyword ranked search over encrypted cloud data." In 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), pp. 497-503. IEEE, 2017.
6. Deepa, N., Pandi Vijayakumar, Bharat S. Rawal, and B. Balamurugan. "An extensive review and possible attack on the privacy preserving ranked multi-keyword search for multiple data owners in cloud computing." In 2017 IEEE International Conference on Smart Cloud (SmartCloud), pp. 149-154. IEEE, 2017.
7. Fu, Zhangjie, Xingming Sun, Zhihua Xia, Lu Zhou, and Jiangang Shu. "Multi-keyword ranked search supporting synonym query over encrypted data in cloud computing." In 2013 IEEE 32nd International Performance Computing and Communications Conference (IPCCC), pp. 1-8. IEEE, 2013.
8. Peng, Tianyue, Yaping Lin, Xin Yao, and Wei Zhang. "An efficient ranked multi-keyword search for multiple data owners over encrypted cloud data." IEEE Access 6 (2018): 21924-21933.
9. Zhang, Wei, Yaping Lin, Sheng Xiao, Jie Wu, and Siwang Zhou. "Privacy preserving ranked multi-keyword search for multiple data owners in cloud computing." IEEE Transactions on Computers 65, no. 5 (2015): 1566-1577.
10. Dai, Hua, Yan Ji, Geng Yang, Haiping Huang, and Xun Yi. "A Privacy-preserving Multi-keyword Ranked Search over Encrypted Data in Hybrid Clouds." IEEE Access (2019).
11. Zhao, Ruihui, Hongwei Li, Yi Yang, and Yu Liang. "Privacy-preserving personalized search over encrypted cloud data supporting multi-keyword ranking." In 2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1-6. IEEE, 2014.
12. Ajai, Ajni K., and R. S. Rajesh. "Hierarchical Multi-Keyword Ranked search for secured document retrieval in public clouds." In 2014 International Conference on Communication and Network Technologies, pp. 33-37. IEEE, 2014.
13. Chen, Chi, Xiaojie Zhu, Peisong Shen, Jiankun Hu, Song Guo, Zahir Tari, and Albert Y. Zomaya. "An efficient privacy-preserving ranked keyword search method." IEEE Transactions on Parallel and Distributed Systems 27, no. 4 (2015): 951-963.
14. Xu, Jian, Xinyu Huang, Geng Yang, and Yuanyuan Wu. "An Efficient Multi-keyword top-k Search Scheme over Encrypted Cloud Data." In 2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN), pp. 305-310. IEEE, 2018.
15. Mlgheit, Jassim R., Essam H. Houssein, and Hala H. Zayed. "Efficient Privacy Preserving of Multi-keyword Ranked Search Model over Encrypted Cloud Computing." In 2018 1st International Conference on Computer Applications & Information Security (ICCAIS), pp. 1-6. IEEE, 2018.
16. Wang, J., Ma, H., Tang, Q., Li, J., Zhu, H., Ma, S., & Chen, X. "A new efficient verifiable fuzzy keyword search scheme". In 2012 Journal of Wireless Mobile Networks, Ubiquitous Computing and Dependable Applications, 3(4), pp. 61-71.