A SECURE AND DYNAMIC MULTI KEYWORD RANKED SEARCH SCHEME FOR CLOUD DATA

1. A.Hemanth Kumar, 2.M.Rajesh
1.Associate Professor,Dept of MCA, Audisankara Institute of Technology,(AUTONOMOUS),Gudur,AP,India.
2.PG Scholar,Dept of MCA, Audisankara Institute of Technology Gudur,AP, India.

Abstract – With the explosive growth of data volume in the cloud computing environment, data owners are increasingly inclined to store their data on the cloud. Although data outsourcing reduces computation and storage costs for them, it inevitably brings new security and privacy concerns, as the data owners lose direct control of sensitive data. Meanwhile, most of the existing ranked keyword search schemes mainly focus on enriching search efficiency or functionality, but lack of providing efficient access control and formal security analysis simultaneously. To address these limitations, in this paper we propose an efficient and privacy-preserving Multi-keyword Ranked Search scheme with Fine-grained access control (MRSF). MRSF can realize highly accurate ciphertext retrieval by combining coordinate matching with Term Frequency-Inverse Document Frequency (TF-IDF) and improving the secure kNN method.

Index terms – Ranked Search, Multi Keyword Search, Cloud Computing, Access control.

I. INTRODUCTION

AS a new computing paradigm [1], cloud computing offers ubiquitous and on-demand access to flexible computation and storage resources. Therefore, outsourcing local data to cloud servers has become a common practice for enterprises and individuals. While this measure greatly reduces hardware and maintenance expenditure, data owners actually lose direct control over their data. This certainly has brought some security concerns, especially to owners of highly sensitive data (i.e., electronic medical records, financial documents, etc.). With such suspicion, individuals and enterprises may be reluctant to outsource their sensitive data to an untrusted third-party cloud service provider. Thus, security concerns will become one of the primary obstacles impeding the widespread deployments of cloud computing. To prevent potential data leakage, data owners usually encrypt their data before outsourcing them to the commercial public cloud. However,

conventional data encryption schemes disable the cloud from running authorized calculations on its storage (e.g., retrieving the interested file for a certain customer), which disables the implementation of plaintext-based information retrieval technologies over outsourced data. A trivial solution is to download all the data and decrypt them locally, but this may lead to a huge waste of bandwidth and computation resources. Thus, how to achieve efficient data retrieval while ensuring data security becomes a challenging issue. The Searchable Symmetric Encryption (SSE) is broadly considered as a promising way to solve the dilemma between data utilization and confidentiality. Some inspiring SSE-based designs include Boolean keyword search schemes in these schemes enable conjunctive keyword search over encrypted data. However, none of these schemes are adequate to provide a ranked search. The complicated design of SSE also prohibits its direct application in large-scale cloud data. To address the former issue, the first secure ranked search scheme is proposed in, but it just supports single keyword search.

II. BACK GROUND WORK

Searchable encryption is a promising technique to replace the trivial method that the user downloads encrypted outsourced data then decrypts it to search. The existing works mainly contribute in two ways: the encryption structure and the expansion of search functionalities. In this section, we review some recent achievements in this area in two aspects. Searchable Encryption. SE schemes can be divided into two categories, namely, Asymmetric Searchable Encryption (ASE) and Symmetric Searchable Encryption (SSE). The pioneering work proposed by Boneh et al. in is the first public-key encryption scheme that supports single keyword search. This work is extended in supporting more operations over encrypted data such as conjunctive keyword search, range query, etc. However, the ASE schemes are less efficient than SSE schemes due to the complex encryption procedures. Yu et al. proposed a two-round searchable encryption (TRSE) scheme that supports ranked multi-keyword search. In TRSE, homomorphic encryption is leveraged to encrypt index and query generated by a vector space model. Although TRSE guarantees high security, it takes two rounds of communications between the data user and the cloud server to complete one search process. In the work of Cheng et al., a public-key cryptosystem based kNN scheme is proposed. Different from the former secure kNN methods based on symmetric encryption, the proposed scheme leverages the distributed two trapdoors public-key cryptosystem (DT-

PKC), which enables secure k-NN query with multiple keys. The concept of SSE is first proposed by Song et al. in, but this scheme lacks support for keywords relevance calculation and multi-keyword search. Another SSE based ranked search over encrypted data is proposed by Wang et al. in, leveraging OPSE (order-preserving symmetric encryption). In order to conduct kNN search over encrypted dataset, Wong et al. first proposed the asymmetric scalar product-preserving encryption (ASPE) in, which is viewed as the original secure kNN scheme. Since then, ASPE has been thoroughly studied in many works ,However, most ranked keyword search schemes based on ASPE are vulnerable to level-3 attack, where the adversary is able to gain a certain amount of plaintext ciphertext pairs. In a word, secure kNN computation is anSE method with high usability but relatively low security. Functionality extension. Proposed schemes in literature support at least one search functionality. As mentioned before, focus on boolean keyword search, while, focus on multi-keyword search. Schemes that focus on geometric search include. Relevance scores in keyword/textual search are replaced by the distance between coordinate points, and the data structure is often designed specially in those schemes. Other schemes support mixed search objects, for example, proposed a scheme that returns top-k location points with keywords matching the queried keywords. In this paper, we mainly focus on keyword search over encrypted data. There are variant functionality extensions towards keyword search schemes over encrypted data.

Towards semantic-aware keyword search over encrypted data, many works provide a solution. Guan et al. proposed a multi-keyword ranked search scheme with a semantic extension (CLRSE) that can be applied toa cross-lingual dataset. The data user is allowed to set the language preference before launching the search. CLRSE adopts a two-cloud system model and Paillier cryptosystem to achieve a higher security level. To better extract search intents of the data user, Dai et al. proposed a scheme that enables semantic-aware keyword search over encrypted data. They adopt the secure kNN method as the encryption algorithm. This scheme utilizes a natural language processing model to extract features from both the document and search query, such that the document-query similarity is transformed into feature vectors similarity. Lang et al. adopted the same natural language processing model in their semantic-based compound secure keyword search scheme, they also utilize Locality-Sensitive Hashing (LSH)to eliminate

unnecessary privacy leakage. To improve users' query experiences and further enrich search functionality, more attention is devoted to conjunctive and disjunctive secure searches. In, Yin et al. proposed a multi-keyword conjunctive secure query scheme based on the multi-owner system architecture. The data user generates a trapdoor set to launch the conjunctive query. During the search process, the cloud server builds an equation for each document and checks the correctness of the equation to judge if the document satisfies the query. The document indexes are encrypted with bilinear pair map, which involves considerable computation. Li et al. proposed a fine-grained multi-keyword search (FMS) scheme over encrypted cloud data, which also leverages the TF-IDF rule to calculate relevance scores. FMS is a variant of the secure kNN method with more comprehensive functionality. It supports dynamic adjustment to data users' preference factors and conjunctive operations in the multi-keyword search. Dynamic updating is also a practical technique needed in many application scenarios. In, Xu et al. presented a multi-keyword scheme that supports efficient updates for keyword dictionary. When a new keyword is added into the keyword dictionary, the data owner does not need to reconstruct the whole dictionary. Li et al. proposed a dynamic searchable encryption scheme (SEPSSE) by combining secure kNN method and Attribute-Based Encryption(ABE) together. A trusted authority is contained in the system model of SEPSSE, which generates the ABE keyfor document encryption. To realize secure data insertion and deletion in SEPSSE, the authors integrate the updating operations with searchable encryption, so that the updating process can be performed with forward and backward privacy. Because the cloud server is deemed as not fully trusted, many works are proposed to verify the correctness of the search results. As in, Ge et al. proposed a keyword search scheme with symmetric-key based verification, which is able to verify the correctness of the search results. Since the verification scheme is built upon the ingenious Accumulative Authentication Tag (AAT), it avoids complex operations. In, Sun et al. proposed a privacy preserving multi-keyword text search scheme (BMTS) that supports similarity-based ranking and result verification. To improve the search accuracy, BMTS adopts the TF-IDF rule to evaluate keyword weights and uses cosine measure as the similarity evaluation function. Miao et al. built a basic Verifiable SE Framework against insider Keyword-Guessing Attack (KGA) by extending the public auditing technique to the SE scheme.

Through extending the framework, the enhanced scheme can support multi-keyword search, multi key encryption, and dynamic update.The existing works over access control mostly focus on complicated methods such as Cipher-Policy Attribute-Based Encryption (CP-ABE) and the Cipher-Policy Attribute-Based Keyword Search (CP-ABKS). CP-ABE schemes incur heavy computational costs, which often grow with the complexity of the access structure. The computational and storage costs of existing CP-ABKS schemes are approximately proportional to the complexity of access policy as well. As it is concluded in, approaches that adopt asymmetric encryption methods to protect data from unauthorized access usually have huge key numbers, which incurs a high key management burden. Thus, inventing a light-weight access control mechanism over keyword search schemes is still indemand.

III. PROPOSED WORK

In this paper, we consider a cloud storage system that supports ranked document retrieval in a privacy-preserving way. As illustrated in Fig. 1, we consider three basic entities in our system model, namely the data owner, the cloud server, and the data user. The data owner ought to submit his/her encrypted data documents to the cloud server. Before data outsourcing, the data owner first builds encrypted searchable indexes for all data documents, then sends both indexes and encrypted documents to the cloud. Besides, the data owner decides the access roles for different data users. The cloud server, which has exceptional computation power and huge storage capacities, provides data hosting and processing services for data owners and data users. Upon receiving the token from an authorized data user, the cloud server first conducts search operations based on encrypted indexes and token, then returns the relevant encrypted documents. The data user acquires the secret keys and the access roles from the data owner through a secure channel after issuing a search request. Next, the data user generates his/her search token with the secret key, then sends it to the cloud server. The secret key is also used for decrypting the retrieved results off-line. Moreover, the polynomial based access control mechanism is employed to manage the decryption capabilities of data users.

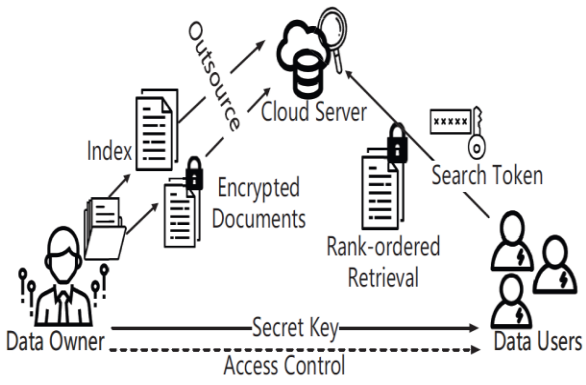

Fig. 1: System Overview

Implementation Modules

- **Data Owner**
- The data owner ought to submit his/her encrypted data documents to the cloud server. Before data outsourcing, the data owner first builds encrypted searchable indexes for all data documents, then sends both indexes and encrypted documents to the cloud. Besides, the data owner decides the access roles for different data users.
- **Cloud Server**
- The cloud server, which has exceptional computation power and huge storage capacities, provides data hosting and processing services for data owners and data users.
- Upon receiving the token from an authorized data user, the cloud server first conducts search operations based on encrypted indexes and token, then returns the relevant encrypted documents.
- **Data User**
- The data user acquires the secret keys and the access roles from the data owner through a secure channel after issuing a search request.
- Next, the data user generates his/her search token with the secret key, then sends it to the cloud server.
- The secret key is also used for decrypting the retrieved results off-line.

Implementation Algorithm

Algorithm 2: Search

Input: Index $\widehat{D}$ and encrypted document set C from the data owner; Token $\widehat{Q}$ from the data user;
Output: Ranked namelist of top-k relevance scores and their corresponding documents;

for i *from* 1 *to* n **do**
  $\Psi_i \leftarrow \Psi(\widehat{D}_i, \widehat{Q})$;
  if $|\Psi_i| \gg \alpha(\widehat{N})$ **then**
    delete Ψ_i;
  else
    $Result \leftarrow Result.append(\Psi_i)$;
... // Rank the elements in $Result$.
for i *from* 1 *to* k **do**
  $F_{\widetilde{\mathcal{W}}} \leftarrow F_{\widetilde{\mathcal{W}}}$.append($Result$[i].name);
for i *from* 1 *to* n **do**
  if C_i.name in $F_{\widetilde{\mathcal{W}}}$ == true **then**
    $C_{\widetilde{\mathcal{W}}}$.append($C_i$);
return $F_{\widetilde{\mathcal{W}}}$, $C_{\widetilde{\mathcal{W}}}$.

IV. RESULTS

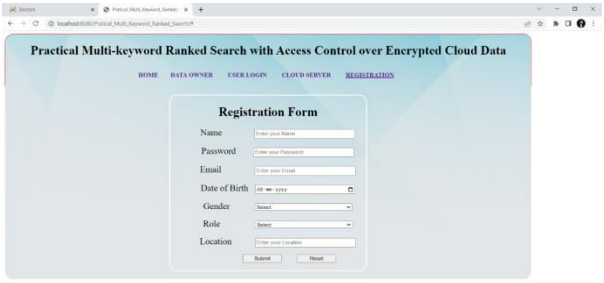

Fig. 2: Data Owner Registration

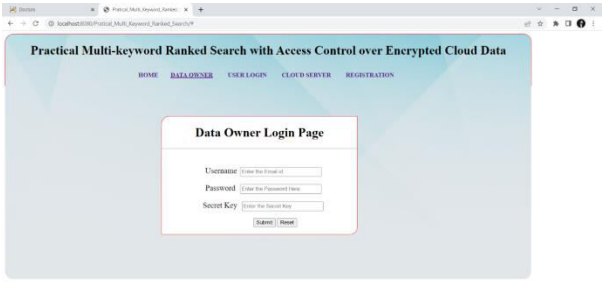

Fig. 3: Data Owner login

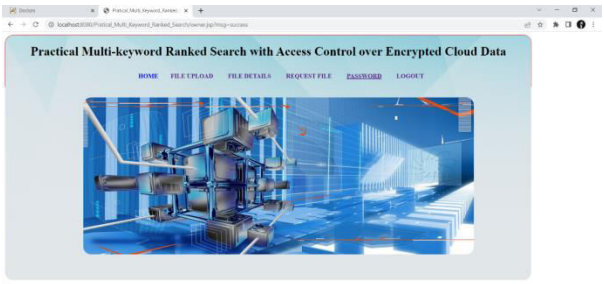

Fig. 4: Data owner Home

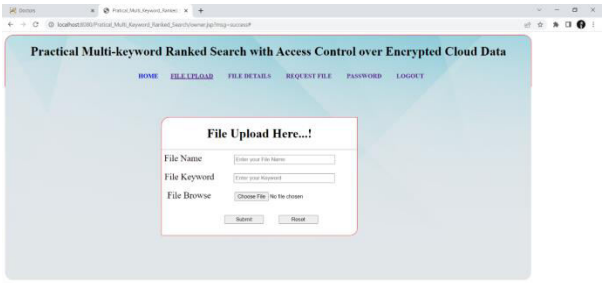

Fig. 5: File Upload

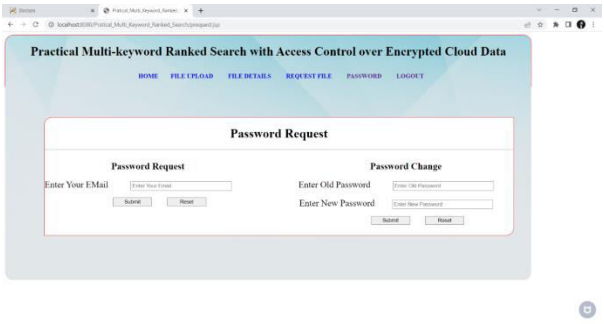

Fig. 6: Change Password.

V. CONCLUSION

In this paper, we propose a privacy-preserving multi keyword search scheme with lightweight fine-grained access control (MRSF). Compared with previous schemes, besides realizing access control, MRSF achieves a better search performance and higher security level. In order to improve the practicability and security of MRSF, we combine the TF-IDF rule with the conventional coordinate matching method and integrate the access control strategy with the improved secure kNN scheme. Formal security definitions and corresponding analysis show that MRSF is IND-CLS-CPA secure, we also prove that MRSF is resistant to the representative KPAs. Finally, extensive evaluations demonstrate the influential factors for search accuracy and efficiency of MRSF.

REFERENCES

[1] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, “Secure rankedkeyword search over encrypted cloud data,” in IEEE InternationalConference on Distributed Computing Systems, 2010.

[2] L. Zhang, Y. Zhang, and H. Ma, “Privacy-preserving and dynamicmulti-attribute conjunctive keyword search over encrypted clouddata,” IEEE Access, vol. 6, pp. 34 214–34 225, 2018.

[3] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-preservingmulti-keyword ranked search over encrypted cloud data,” IEEETransactions on Parallel and Distributed Systems, vol. 25, no. 1, pp.222–233, Jan 2014.

[4] D. X. Song, D. Wagner, and A. Perrig, “Practical techniques forsearches on

encrypted data," in Proceeding 2000 IEEE Symposiumon Security and Privacy. S P 2000, May 2000, pp. 44–55.

[5] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Applied Cryptography andNetwork Security. Berlin, Heidelberg: Springer Berlin Heidelberg,2005, pp. 442–455.

[6] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," Journal of Computer Security, vol. 19, no. 5, pp. 895–934,2011.

[7] M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J. Malone-Lee, G. Neven, P. Paillier, and H. Shi, "Searchable encryption revisited:consistency properties, relation to anonymousibe, and extensions," in Advances in Cryptology – CRYPTO 2005.Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 205–222.

[8] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advances in Cryptology -EUROCRYPT 2004, C. Cachin and J. L. Camenisch, Eds. SpringerBerlin Heidelberg, 2004, pp. 506–522.

[9] M. Bellare, A. Boldyreva, and A. O'Neill, "Deterministic and efficiently searchable encryption," in Advances in Cryptology - CRYPTO 2007. Berlin, Heidelberg: Springer Berlin Heidelberg,2007, pp. 535–552.

[10] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized private keywordsearch over encrypted data in cloud computing," in 2011 31stInternational Conference on Distributed Computing Systems. IEEE,2011, pp. 383–392.

AUTHORS PROFILES:

A.HEMANTHA KUMAR has received his M.Tech degree in CSE from Sathyabama Deemed University in 2006, Chennai.He is dedicated to teaching field from 2001. He has guided P.G and U.G students. His research areas included Computer Networks, Network Security and Machine Learning. At present he is working as Associate Professor in Audisankara Collegeof Engineering and Technology,Gudur, Tirupati(Dt), Andhra Pradesh, India.

MULI RAJESH has Pursuing his MCA from Audisankara InstituteOftechnology(AUTONOMOUS),Gudur affiliated to JNTUA in 2022.Andhra Pradesh,india.