An Efficient and Secure Cipher-Text Retrieval Scheme Based on Mixed Homomorphic Encryption and Multi-Attribute Sorting Method Under Cloud Environment

Lin Teng, Hang Li, Jie Liu, and Shoulin Yin (Corresponding author: Hang Li)

Software College, Shenyang Normal University No.253, HuangHe Bei Street, HuangGu District, Shenyang, P.C 110034, China (Email: 1451541@qq.com) (Received May 11, 2017; revised and accepted Aug. 26, 2017)

Abstract

Cipher-text retrieval plays an important role in data encryption storage service under cloud environment. The present sorting search algorithms have low precision due to corresponding score computing with single local attribute, which cannot accurately sort according to the similarity. To solve this problem, we propose an efficient and secure cipher-text retrieval scheme based on mixed homomorphic encryption and multi-attribute sorting method under cloud environment. This new scheme is divided into four steps: 1) Constructing multiple attribute characteristic vector safety index for documents; 2) Constructing reverse index for the uploaded documents and generating vector set of document, then computing module of each document vector; 3) Encrypting document vector set with homomorphic encryption and uploading them into cloud; 4) Adopting multiple attribute score formula to calculate document relevance score, according to the scores ranking to return the interesting retrieval results for users. Experiments show that our method has the higher retrieval speed and the better retrieval efficiency under cloud environment.

Keywords: Cipher-Text Retrieval; Cloud Environment; Homomorphic Encryption; Multi-Attribute Sorting Method; Relevance Score

1 Introduction

Cloud computing [24, 25] is the comprehensive development of parallel computing, distributed computing and grid computing. Cloud has attracted widespread attention and recognition as it transfers the traditional computing and storage functions into the cloud environment, which saves lots of hardware cost for users. Currently,

the typical cloud platforms are EC2 [3], Google App Engine [11] and Microsoft Live Mesh. With the development of cloud, more sensitive information (such as medical records, financial information and important documents of company) are stored in cloud [1, 2]. Once the data are received by cloud provider, users lose the directly control for their data, which can cause the leak of privacy data. Encryption is an effective method to protect privacy of users' data. However, this way loses many features and can lead to difficult encryption [5, 6, 9, 21]. Especially, how to conduct encrypted data query in untrusted cloud environment has aroused people's attention.

Currently, most ciphertext retrieval schemes do not support sort search. Especially, under cloud computing environment with a large scale of data, there may be lots of documents including one keyword [15, 18]. How to find the closest document in some documents is difficulty. Many researchers had proposed a lot of schemes to improve it. Tan [26] proposed a non-circuit based Ciphertext Policy-Attribute Based Homomorphic Encryption scheme to support outsourced cloud data computations with a fine-grained access control under the multiuser scenario. First, he incorporated Attribute Based Encryption scheme into homomorphic encryption scheme in order to provide a fine-grained access control on encrypted data computation and storage. Then, the proposed scheme was further extended into non-circuit based approach in order to increase the practical efficiency between enterprise and cloud service providers. The results had greatly reduced the computation time and ciphertext size. But the new scheme had a low retrieval efficiency. Gong [12] presented an encryption scheme, based on the composite degree residuosity classes, which could block chosen ciphertext attack while maintaining homomorphism. Elhoseny [10] proposed a novel encryption schema based on Elliptic Curve Cryptography and

homomorphic encryption to secure data transmission in WSN. To reduce energy consumption of cluster head, homomorphic encryption was used to allow cluster head to aggregate the encrypted data without having to decrypt them. He demonstrated that the proposed method was capable to work with different sensing environments that needed to capture text data as well as images. Lu [22] proposed encryption of all genotype and phenotype data to maintain the privacy of subjects. Chen [8] analyzed the secrecy capacity between the source station and the destination station based on the Shannon third theorem on channel capacity, which introduced some secure cooperative communication system. Hou [14] proposed a robust remote authentication scheme with privacy protection, which achieved the efficiency. Arup [7] proposed scheme by using simple Boolean based encryption and decryption of the data files, which was low in computational cost. Zhang [29] presented a pairing-based multi-user homomorphic encryption scheme to privately outsource computation of different users. And there are many other methods, such as [19, 20, 28]. But retrieval efficiency is still low in their methods.

Therefore, this paper proposes an efficient and secure cipher-text retrieval scheme based on mixed homomorphic encryption and multi-attribute sorting method. This new scheme can execute distinction sorting according to the different sorting ways selected by the user, and return the most suitable retrieval results. From the keywords local property and global property, multi-attribute sorting comprehensively reflects the document characteristics, which not only fully considers the difference under the same keywords in different documents, but ensures the documents' quality and authority. Meanwhile, it adopts fully homomorphic encryption method that not only ensures the safety of the user's data, but can directly execute addition and multiplication operation for cipher text, this will greatly improve the retrieval efficiency. Compared with the existing schemes, it improves the precision with ensuring data security.

This paper is organized as follows. Section 2 detailed introduces ciphertext retrieval model based on multiattribute sorting and homomorphic encryption. We give the experiment analysis in Section 3. There is a conclusion in Section 4.

2 Ciphertext Retrieval Model Based on Multi-Attribute Sorting and Homomorphic Encryption

Data has characteristics with frequent using and a large scale in a cloud computing environment. How to quickly and accurately find the data you need in this environment is a very important problem. Obviously, it is inappropriate to use a linear search algorithm and public key encryption algorithm for retrieval process. In that it

needs much logarithmic operation. Though the existing ciphertext sort search algorithms have small calculation and high speed, the precision is lower. This paper presents a data retrieval model based on multi-attribute sorting to make up for this shortcoming as shown in Figure 1.

From Figure 1, we can know that this new model includes two main processes: document preprocessing and cryptograph retrieval correlation sorting. Document preprocessing is used to extract multi-attribute eigenvector and construct reverse index. Function of cryptograph retrieval correlation sorting is that it adopts multi-attribute scoring function to calculate the document correlation score according to user's retrieval request and return sorting result.

2.1 Determining Keywords

In this step, initial document is as the input. Words will be separated by segmentation method. It will form result set $K = (k_1, \dots, k_n)$ which will be filtered. In this results, it chooses the keywords that can reflect document meaning clearly.

2.2 Multi-Attribute Eigenvector Extraction

Multi-attribute eigenvector extraction is a key step for document preprocessing. It can extract the multiattribute eigenvector of keywords. The followings are some definitions.

Definition 1. Keyword local property. The contribution of keyword for document can be affected by local factors. These influence factors are defined as keyword local property, such as TFIDF value, word property, word length and word position etc.

Definition 2. Keyword global property. It also can be called document property. The contribution of keyword for document can be affected by global factors. These influence factors are defined as keyword local property, such as document quoted frequency and download number etc.

Definition 3. Local property eigenvector. *Eigenvector* using keyword local property can be called local property eigenvector.

Definition 4. Global property eigenvector. *Eigenvector* using keyword global property can be called global property eigenvector.

Definition 5. Single property eigenvector. If document eigenvector only uses one keyword property that it can be called single property eigenvector.

Definition 6. Multi-property eigenvector. It consists of keyword local property and global property as Figure 2.

The *keywords* are extracted from document. *Attributes* indicate local and global properties.



Figure 1: Ciphertext retrieval model based on multi-attribute sorting



Figure 2: Document multi-attribute eigenvector

2.3 Homomorphic Encryption

Homomorphism is the concept of modern algebra [23]. Supposing $\langle G, \cdot \rangle$ and $\langle H, * \rangle$ are two algebra systems. If $\forall a, b \in G$, then $f(a \cdot b) = f(a) * f(b)$. The f can be called homomorphic mapping from G to H. Cryptography promotes the concept of homomorphic mapping in modern algebra. Encryption is a mapping from the plaintext to cryptographic [4]. And if the encryption mapping is a homomorphic mapping, we can say that it is a homomorphic encryption scheme. In this paper, homomorphic encryption is based on integer modulo arithmetic, its processes are described as follows:

- Kengen. Randomly select a P-digit big prime number as key p.
- Encryption. Randomly select a Q digit big prime number q, and P > Q > plaintextlength. Two random numbers r_1 and r_2 , N = pq. Ciphertext $c = (m + pr_1 + pqr_2) \mod N$.
- Decryption. Plaintext $m = c \mod p$.

Homogeneity analysis. Assuming that two plaintexts m_1 and m_2 , and corresponding ciphertexts c_1 and c_2 . So

$$c_1 = (m_1 + pr_{11} + pqr_{12}) \mod N.$$

$$c_2 = (m_2 + pr_{21} + pqr_{22}) \mod N.$$

Additive homogeneity analysis.

$$c_1 + c_2 = (m_1 + m_2 + p(r_{11} + r_{21}) + pq(r_{12} + r_{22}) \mod N$$
$$= c(m_1 + m_2).$$

Multiplication homogeneity analysis.

$$c_1 \cdot c_2 = (m_1 \cdot m_2 + pm_1r_{21} + pqm_1r_{22} + pm_2r_{11} + p^2r_{11}r_{21} + p^2qr_{11}r_{22}) \mod N.$$

$$c_1 \cdot c_2 \mod p = m_1 \cdot m_2.$$

In summary, the homomorphic encryption is based on integer modulo arithmetic, which not only meets the additive homogeneity, but meets multiplication homogeneity.

2.4 Reverse Index

Reverse index [16,30] is a data structure used to describe the relation between keywords set and documents set. It stores the storage location mapping of keyword in one document under the full-text retrieval.

The traditional reverse index structure consists of an index file and an reverse file. The index list is a collection of all the keywords in the document. It is composed of all records, each record contains the keywords and keywords' corresponding pointer. Pointer points to the corresponding logical address in the reverse file. The reverse



Figure 3: Multi-attribute reverse index

file indicates that which documents contain this keyword, sion). the frequency information of keywords and addresses set of these documents. Our new reverse index model is as Figure 3. According to multi-attribute eigenvectors, it will change the single attribute keyword frequency in traditional reverse file as multi-attribute including keyword frequency, location, indexed frequency etc.

2.5Score Function of Document

Multi-attribute score function is adopted to execute relevance sorting. Score function can calculate the correlation score between retrieval word and a document.

In our new model, the relevance score computing is according to local property and global property. Due to the different importance degree of attribute, we introduce attribute weight. For example, for attribute $(\rho_1, \rho_2, \cdots, \rho_n)$ of keyword K, its corresponding weight value (o_1, o_1, \cdots, o_n) , and $\sum_{i=1}^{n} o_i = 1$. Therefore, attribute score function of our new scheme can be defined as:

$$score = \sum_{i=1}^{n} \rho_i \times o_i. \tag{1}$$

Where o_i is the weight of attribute ρ_i , which can be dynamic adaptively adjusted according to different sort methods.

3 **Experience** Analysis

In order to verify the effectiveness of new scheme, we make comparison experiments under MATLAB environment.

First, we select evaluation criteria: Recall rate (RE), Precision rate(PE) and MAP (Mean of Average Preci-

$$RE = \frac{SRE}{TRE}.$$
 (2)

$$PE = \frac{SRE}{TNE}.$$
 (3)

$$MAP = \sum_{i}^{r} \frac{1}{r}.$$
 (4)

Where SRE and TRE denote system retrieved relevant files and total number of relevant files respectively. TNEis the total number of retrieved files. If the retrieved file is closer to the top, the MAP is likely to rank higher.

Second, we select 50 papers from Google Scholar and set global attribute for all papers. When constructing retrieve model, each paper will be preprocessed. We set title, content, keywords and summary as retrieve information. CP-ABHER-LWE [27], HACC [17] and ECCH [13] are compared with our method (abbreviated to HEMAS) to show the better results of our scheme under the same experiment environment.

It conducts same query in single attribute cryptograph retrieval system and multi-attribute cryptograph retrieval system and returns to the results of top 20, 30, 40. Table 1 presents the results. It can be seen that our method is better than other three schemes.

Table 1: Experimental evaluation results

Index	CP-ABHER-LWE	HACC	ECCH	HEMAS
RE	93.4%	95.8%	94.5%	96.8%
PE	94.6%	96.2%	95.5%	98.4%
MAP	94.9%	96.1%	94.9%	97.5%

Similarly, it conducts same query in single attribute cryptograph retrieval system and multi-attribute cryptograph retrieval system and returns to the results of top 10, 20, 30, 40. Table 2 is the retrieval time comparison.



Figure 4: Precision comparison with new scheme

From the table, the retrieval time is greatly reduced with new scheme.

 Table 2: Retrieval time

	CP-ABHER			
Index	-LWE	HACC	ECCH	HEMAS
Top 10	$188 \mathrm{ms}$	190ms	$205 \mathrm{ms}$	$163 \mathrm{ms}$
Top 20	$191 \mathrm{ms}$	$192 \mathrm{ms}$	$208 \mathrm{ms}$	$164 \mathrm{ms}$
Top 30	$193 \mathrm{ms}$	$193 \mathrm{ms}$	$212 \mathrm{ms}$	$166 \mathrm{ms}$
Top 40	195 ms	$195 \mathrm{ms}$	$214 \mathrm{ms}$	$168 \mathrm{ms}$

We only make experiment with our new method using different attributes: one attribute, two attributes, four attributes as shown in Figure 4.

We can see that precision will become more higher with the increasing of attribute. In summary, this paper proposes the efficient and secure cipher-text retrieval scheme based on mixed homomorphic encryption and multi-attribute sorting method under cloud environment, which is effectiveness for improving retrieval accuracy rate.

3.1 Performance Analysis

We also make a comparison to CP-ABHER-LWE [27], HACC [17] and ECCH [13] with our HEMAS method. The following is the explanation of symbols in this section: p: bilinear operation. e: exponent operation. s: point multiplication operation in G_1 . $|G_1|$: the element length of corresponding group. |m|: the length of message. |U|: the length of user identity.

Table 3 shows the calculation about the four algorithms. And we can know that users with HEMAS only need one pairing operation and one point multiplication operation in Homomorphic encryption stage less than CP-ABHER-LWE, HACC and ECCH. In Reverse index stage, HEMAS needs 2n - 1 pairing operations obviously superior to ECCH. The pairing operation number is more than CP-ABHER-LWE, HACC and ECCH. In that our new scheme dose not need exponent operation, the total calculation is superior to CP-ABHER-LWE, HACC and ECCH when n is big.

In order to specifically analyze running time, we use the A type elliptic curve to test in jpbc database. Then we record the running time with the above schemes as table4 from MATLAB platform.

Table 4 shows that the running time with our new scheme is less than other schemes. It is the optimal scheme.

4 Conclusions

We present a mixed retrieval scheme in this paper, which shows the following merits.

- 1) Fast retrieval speed. The new scheme only needs to search the safety index list that is very suitable for the data with a large scale in the cloud environment.
- 2) Less calculation. Due to homomorphic encryption, it greatly decreases the calculation time.
- 3) **Precision improved**. This scheme introduces multi-attribute eigenvectors including keywords global properties and local properties, which makes comprehensive evaluation for documents from several aspects, so as to return the most relevant documents.
- 4) Shorten retrieval time. In that the precision of score computing greatly reduces the sorting time, so retrieval time is reduced too.

References

- D. S. AbdElminaam, "Improving the security of cloud computing by building new hybrid cryptography algorithms," *International Journal of Electronics and Information Engineering*, vol. 8, no. 1, pp. 40-48, 2018.
- [2] S. Almulla, Y. Y. Chan, "New secure storage architecture for cloud computing," *Future Information Technology*, pp. 75-84, 2011.
- [3] B. Amazon, "Amazon elastic compute cloud (Amazon EC2)," Virtual Grid Computing, 2010.
- [4] Z. Cao, L. Liu, Y. Li, "Ruminations on Fully Homomorphic Encryption in Client-server Computing Scenario," *International Journal of Electronics and Information Engineering*, vol. 8, no. 1, pp. 32–39, 2018.
- [5] T. Y. Chang and M. S. Hwang, W. P. Yang, "An improved multi-stage secret sharing scheme based on the factorization problem," *Information Technology* and Control, vol. 40, no. 3, pp. 246-251, 2011.
- [6] T. Y. Chang, M. S. Hwang, W. P. Yang, K. C. Tsou, "A modified ohta-okamoto digital signature for batch verification and its multi-signature version," *International Journal of Engineering and Industries* (*IJEI*'12), vol. 3, no. 3, pp. 75-83, Sep. 2012.

Index	CP-ABHER-LWE	HACC	ECCH	HEMAS
Homomorphic encryption	3ne+s	n(2p+s+3e)	3n(p+2e)	n(s+p)
Reverse index	4ns - 2s + 2np + p	n(2p+s)	n(2e+3p)	(2n-1)p

Table 3.	Calculation	comparison	with	different	schemes
Table 9.	Calculation	comparison	W1011	unicient	sentenies

Index	CP-ABHER-LWE	HACC	ECCH	HEMAS
Homomorphic encryption	258.647 ms	241.323ms	$109.752 \mathrm{ms}$	50.219ms
Reverse index	$352.458 \mathrm{ms}$	$288.943 \mathrm{ms}$	$209.546\mathrm{ms}$	$68.252 \mathrm{ms}$

Table 4: Calculation time comparison with different schemes

- [7] A. K. Chattopadhyay, A. Nag, K. Majumder, "Secure data outsourcing on cloud using secret sharing scheme," *International Journal of Network Security*, vol. 19, no. 6, pp. 912, 2017.
- [8] J. S. Chen, C. Y. Yang, M. S. Hwang, "The capacity analysis in the secure cooperative communication system," *International Journal of Network Security*, vol. 19, no. 6, pp. 863-869, 2017.
- [9] S. F. Chiou, M. S. Hwang, S. K. Chong, "A simple and secure key agreement protocol to integrate a key distribution procedure into the DSS," *International Journal of Advancements in Computing Technology* (*IJACT'12*), vol. 4, no. 19, pp. 529-535, Oct. 2012.
- [10] M. Elhoseny, H. Elminir, A. Riad, et al., "A secure data routing schema for WSN using elliptic curve cryptography and homomorphic encryption," Journal of King Saud University - Computer and Information Sciences, vol 28, no. 3, pp. 262-275, 2015.
- [11] B. Ferriman, T. Hamed, Q. H. Mahmoud, "Storming the cloud: A look at denial of service in the google app engine," in *IEEE International Conference on Computing, Networking and Communications*, pp. 363-368, 2015.
- [12] L. Gong, S. Li, Q. Mao, et al. "A homomorphic encryption scheme with adaptive chosen ciphertext security but without random oracle," *Theoretical Computer Science*, vol. 609, pp. 253-261, 2016.
- [13] M. Q. Hong, P. Y. Wang, W. B. Zhao, "Homomorphic encryption scheme based on elliptic curve cryptography for privacy protection of cloud computing," in *IEEE International Conference on Big Data Security on Cloud*, pp. 152-157, 2016.
- [14] G. Hou, Z. Wang, "A robust and efficient remote authentication scheme from elliptic curve cryptosystem," *International Journal of Network Security*, vol. 19, no. 6, pp. 904-911, 2017.
- [15] S. T. Hsu, C. C. Yang, and M. S. Hwang, "A study of public key encryption with keyword search", *International Journal of Network Security*, vol. 15, no. 2, pp. 71-79, Mar. 2013.
- [16] M. S. Hwang and C. C. Lee, "Research issues and challenges for multiple digital signatures," *International Journal of Network Security*, vol. 1, no. 1, pp. 1-7, 2005.

- [17] C. Lan, H. Li, C. Wang, "A new key-aggregate encryption scheme with chosen ciphertext security," in *IEEE Sixth International Conference on Information Science and Technology*, pp. 229-233, 2016.
- [18] C. C. Lee, Shih-Ting Hsu, M. S. Hwang, "A study of conjunctive keyword searchable schemes," *International Journal of Network Security*, vol. 15, no. 5, pp. 311–320, 2013.
- [19] H. Li, S. L. Yin, Chu Zhao and Lin Teng, "A proxy re-encryption scheme based on elliptic curve group," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 1, pp. 218-227, Jan. 2017.
- [20] J. Liu, S. L. Yin, H. Li and L. Teng, "A density-based clustering method for K-anonymity privacy protection," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 1, pp. 12-18, Jan. 2017.
- [21] L. Liu, W. Kong, Z. Cao, J. Wang, "Analysis of one certificateless encryption for secure data sharing in public clouds," *International Journal of Electronics* and Information Engineering, vol. 6, no. 2, pp. 110-115, 2017.
- [22] W. J. Lu , Y. Yamada, J. Sakuma "Privacypreserving genome-wide association studies on cloud environment using fully homomorphic encryption," *BMC Medical Informatics and Decision Making*, vol. 15, 2015.
- [23] M. Ogburn, C. Turner, P. Dahal, "Homomorphic encryption," *Proceedia Computer Science*, vol. 20, pp. 502-509, 2013.
- [24] Z. Qingchen, L. T. Yang, C. Zhikui, "Privacy preserving deep computation model on cloud for big data feature learning," *IEEE Transactions on Computers*, vol. 65, no. 5, 2016.
- [25] Z. Qingchen, Z. Hua, L. T. Yang, C. Zhikui, B. Fanyu, "PPHOCFS: Privacy preserving high-order CFS algorithm on cloud for clustering multimedia data," ACM Transactions on Multimedia Computing, Communications and Applications, 2016. DOI: 10.1145/2886779.
- [26] S. F. Tan, A. Samsudin, "Ciphertext policy-attribute based homomorphic encryption (CP-ABHER-LWE) scheme: A fine-grained access control on outsourced cloud data computation," *Journal of Information*

2017.

- [27] S. F. Tan, A. Samsudin, "Ciphertext policy-attribute based homomorphic encryption (CP-ABHER-LWE) scheme: A fine-grained access control on outsourced cloud data computation," Journal of Information Science & Engineering, vol. 33, no. 3, pp. 675-694, 2017.
- [28] S. L. Yin and J. Liu, "A K-means approach for mapreduce model and social network privacy protection," Journal of Information Hiding and Multimedia Signal Processing, vol. 7, no. 6, pp. 1215-1221, Nov. 2016.
- [29] W. Zhang, "A pairing-based homomorphic encryption scheme for multi-user settings," International Journal of Technology & Human Interaction, pp. 11, 2016.
- [30] H. Zhu and R. Wang, "Multi-party passwordauthenticated key exchange scheme with privacy preserving using chaotic maps in random oracle model," Journal of Information Hiding and Multimedia Signal Processing, vol. 8, no. 1, pp. 42-53, Jan. 2017.

Biography

Lin Teng received the B.Eng. degree from Shenvang Normal University, Shenyang, Liaoning province, China in 2016. Now, she is a laboratory assistant in Shenyang Normal University. Her research interests include Multimedia Security, Network Security, Filter Algorithm and Data Mining. She had published more than 10 international journal papers on the above research fields. Email:1532554069@qq.com.

Science and Engineering, vol. 33, no. 3, pp. 675-694, Hang Li is a full professor in Software College, Shenyang Normal University. He received his B.S. and M.S. degrees from Northeastern University. He is also a master's supervisor. He has research interests in wireless networks, mobile computing, cloud computing, social networks, network security and quantum cryptography. Professor Li had published more than 30 international journal papers (SCI or EI journals) on the above research fields. Email: 1451541@qq.com.

> Jie Liu is a full professor in Software College, Shenvang Normal University. He received his B.S. and M.S. degrees from Harbin Institute of Technology. He is also a master's supervisor. He has research interests in wireless networks, mobile computing, cloud computing, social networks, network security and quantum cryptography. Professor Liu had published more than 30 international journal papers (SCI or EI journals) on the above research fields. Email: nan127@sohu.com.

> Shoulin Yin received the B.Eng. And M.Eng. degree from Shenyang Normal University, Shenyang, Liaoning province, China in 2013 and 2015 respectively. His research interests include Multimedia Security, Network Security, Filter Algorithm and Data Mining. He received School Class Scholarship in 2015. Yin had published more than 50 international journal papers (SCI or EI journals) on the above research fields. Email:352720214@gg.com.