Future Generation Computer Systems 87 (2018) 251–258

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Image search scheme over encrypted database

Jun Ye [a,b,*], Zheng Xu [c,d], Yong Ding [e]

[a] School of Mathematics and Statistics, Sichuan University of Science & Engineering, Sichuan, China
[b] Guangxi Key Laboratory of Cryptography and Information Security, Guangxi, China
[c] Shanghai University, Shanghai, China
[d] The Third Research Institute of the Ministry of Public Security, Shanghai, China
[e] School of Computer Science and Information Security, Guangxi Key Laboratory of Cryptography and Information Security, Guilin University of Electronic Technology, Guilin, Guangxi, China

HIGHLIGHTS

- A novel request based image search model is proposed.
- A secure search pattern is generated, and a novel matching strategy is used.
- The search accuracy can be controlled by users.
- Low computation cost on the user side.

ARTICLE INFO

Article history:
Received 29 October 2017
Received in revised form 19 December 2017
Accepted 25 February 2018
Available online 7 May 2018

Keywords:
Cloud computing
Image retrieval
Feature vector
Privacy

ABSTRACT

In big data era, too much ordinary information leakage may lead to the leakage of private information. Image search is widely used in many fields. Though there are many studies on image search, most of them are search in plaintext databases. In this paper, we study the image retrieval techniques over encrypted databases. A content-based retrieval scheme for encrypted images is proposed. In the scheme, a blind technique based on discrete logarithm problem is introduced to keep the privacy of feature vectors, and a novel retrieval way is used. This is a flexible scheme, which support fuzzy search. The client can control the search range. And in the search process, the contents of the original image will not be revealed.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Cloud computing, in which a large amount of computing resources and storage space are collected, can provide various computing and storage services for people. Due to the great power of cloud computing, it brings much convenience to the clients. The complex computation tasks can be easily solved by the cloud server. For purpose of saving the local storage space, the client can stores the files in the cloud servers, and retrieve the files when they are needed.

Cloud storage, in which a lot of storage devices are gathered together, is a kind of service-oriented distributed storage system, can provide the storage and management services of massive data [1,2]. Such as space leasing, data storage, backup, sharing and so on. The data storage services are achieved through the application of cluster computing and distributed computation [3,4]. In the system, a large number of different network resources will be used. In order to deal with the files anytime and anywhere, the clients usually store the files in the cloud.

Images are the common files people often use. With the help of the Internet, people can easily share the images for each other. The clients can retrieval the images that they needed from the cloud server conveniently. There are two main techniques for image search, text-based image retrieval (TBIR) and content-based image retrieval (CBIR) [5,6].

TBIR is the traditional image retrieval method. The traditional text retrieval technology was tried to be used in the image retrieval. Thus, the image retrieval is the same as keyword search, which is a kind of matching search process based on keywords. The indices of a image are generated according to the understanding of the image, such as, name, number, size, contents description, image resource, author, time, storage location and so on. The image retrieval is actually becoming the corresponding text retrieval.

CBIR is the search technology based on feature vectors. The image is analyzed by software, and the content information is

* Corresponding author at: School of Mathematics and Statistics, Sichuan University of Science & Engineering, Sichuan, China.
E-mail addresses: yejun@suse.edu.cn (J. Ye), Xuzheng@shu.edu.cn (Z. Xu), stone_dingy@126.com (Y. Ding).

https://doi.org/10.1016/j.future.2018.02.045
0167-739X/© 2018 Elsevier B.V. All rights reserved.

extracted. The information of color, shape and texture is combined together as the feature vectors, and stored in the feature database. For a given image, in the retrieval process, the feature vectors are extracted, and the similarity between the extracted vectors and the feature vectors stored in the database is computed by using the similarity matching algorithm. The retrieval result is output according to the value of similarity.

The main retrieval principle are the three points. The first one is, form a database retrieval model according to the requirement of the clients. The second one is, collect and process image resources, such as, feature extraction, analysis and indexing. The last one is, use similarity algorithm to retrieval the images with similarity calculation size, index database, the threshold, and output the results according to the similarity descending way.

With the rapid development of the Internet and cloud storage, more and more attention is focused on the study of the retrieval technology. The two main techniques for image search are widely used in modern times. However, in most image search schemes, the images are often stored in plaintext form. Information security is a hot topic in recent years, and people pay more and more attention to protect their private information.

Though the cloud servers are powerful, and their storage space is huge, none of them is completely trusted. In order to protect the privacy and security of the outsourced data, the multimedia data should be encrypted. However, it is difficult to perform operations on the encrypted data. Most of the existing schemes do not support the direct operations on the encrypted multimedia data. The schemes support directly search over the encrypted multimedia data are fairly small. To improve the efficiency in image search, order preserving encryption (OPE) is usually used, which enables people efficiently to identify the real order of data without decrypting the encrypted items. However, Furukawa [7] pointed out if using OPE to encrypt all the numbers in the domain, an attacker can easily obtain the correspondence between the ciphertexts and plaintexts. Therefore, it is significant to study the new effectively retrieve method over the encrypted database.

Our Contributions In this paper, we focus on the retrieval study of encrypted images. Our scheme support directly search on the encrypted multimedia database. And in order to improve the efficiency, the great cost homomorphic encryption is not used. The main contributions of this paper are as follows.

- A novel request based image search model is proposed, which provides a new search way for encrypted images and can overcome some weaknesses of the schemes based on order-preserving encryption.
- The scheme is flexible. The data owner can control the access of the encrypted data. If some one wants to search over the encrypted data, he/she should get the permission of data owner (data owner will return the auxiliary information to the client, with which the search token can be generated). Furthermore, the clients can set the search accuracy (some parameters will be set to control search accuracy, the length and angle of feature vectors).
- A secure search pattern is generated. Clients should get the permission of the data owner, and with the help of data owner, the valid search token of the related image can be generated. When getting the search results, clients have to request data owner for decryption. And the disguise technique is used to hide the outsourced feature vectors and ensure the angle between any two vectors is unchanged.
- A novel matching strategy is used. The comparable encryption is used to compare the length of the encrypted feature vectors, with which the images that are similar to the client's needs will be retrieved easily.
- Low computation cost on the user side. The heavy computation task is outsourced to the cloud server, the computation cost of users is small. And the communication cost is low, only one-round interaction is needed to search for target images between the cloud and the clients.

1.1. Related work

Cloud storage [8–10] has become increasingly prevalent in recent years. It provides a convenient platform for clients to store and share their data in the form of ciphertext. In order to retrieve the encrypted data conveniently, searchable encryption is proposed. It allows a user to securely retrieve the required data.

The first searchable encryption scheme is proposed by Song et al. [11] in 2000. However, the scheme could not resist the statistical analysis attack with multiple queries, and the queried keywords will be revealed. Then, Chang et al. [12] proposed a similar index scheme to improve the efficiency with an encrypted hash table for whole files. The index table is consist of the trapdoor and the information of related file identifiers. Then the formal security notion of searchable encryption is proposed by Curtmola et al. [13].

There also many studies on image search [14–18]. There are two main retrieval methods, TBIR [19,20] and CBIR [21–24]. TBIR is easy to implement and the accuracy is very high. However, there are two main difficulties in text-based retrieval schemes. On one hand, text description is difficult to fully express the rich content of the image because of the limited capacity of description. The different understanding and interest in different areas of the image contents, will lead to the varied establishment. On the other hand, natural language understanding problem has not been solved, it is difficult for computers to read the description based on natural language. The retrieval efficiency is very low when facing the massive database.

In order to improve the search accuracy, content-based retrieval technique is always used. In TBIR system, the information of indices is directly extracted from the image content, and the feature extraction and index generation can be accomplished by computer automatically which greatly improve the retrieval efficiency, such as [25,26]. Chun et al. [27] proposed a content-based image retrieval scheme based on an efficient combination of multi-resolution color and texture features. Chen et al. [28] used clustering algorithm to classify images and improve the retrieval efficiency. Bellafqira et al. [29] proposed a constant-based image search scheme by using homomorphic encryption to retrieval the similar images from the outsourced database. However, the efficiency is not very high. Though image search has been studied for many years, most schemes mainly deals with plaintext images. In 2009, Lu et al. [30] proposed a image search scheme for encrypted images. Cheng et al. [31] used the random projection to generate a content-based image retrieval system. However, there is less research on encrypted image retrieval.

1.2. Organization of this paper

The organization of this paper is as follows. Some preliminaries are given in Section 2. The system model is given in Section 3. Then in Section 4 we introduce the retrieval scheme. The security analysis is given in Section 5. The comparisons and efficiency analysis are given in Section 6. Finally, conclusion will be made in Section 7.

2. Preliminaries

2.1. Hash function

A hash function is an algorithm that takes an arbitrary block of data and returns a fixed-size bit string.

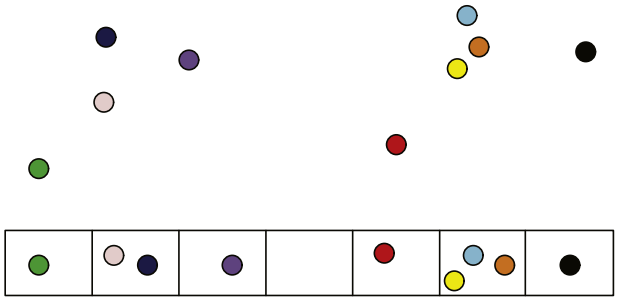

Fig. 1. Locality-sensitive hashing mapping.

The secure hash function has four main properties:

- $\forall x$, there is an efficient computation algorithm to compute $h(x)$;
- $\forall y$, it is computational infeasible to find x, such that $h(x) = y$;
- Given x_1, it is computational infeasible to find x_2, such that $h(x_1) = h(x_2)$;
- It is computational infeasible to find x_1 and x_2, such that $h(x_1) = h(x_2)$.

2.2. Locality-sensitive hashing (LSH)

In many application field, we often face with the very high dimensional data. It is difficult to find a similar (nearest) one from the massive high dimensional data. We can use the hash mapping to transform the high dimensional data fall into the buckets, so that we only need to query the data of the hash map bucket number and search the adjacent data by using the linear matching.

The basic idea of LSH is: two adjacent data points in the original data space over the same projection, a great probability of the two data points are adjacent in the same bucket, but the very small probability of the not adjacent data points.

We define a domain of data points S and a ball for a similarity measure D as $B(\alpha, \gamma) = \{\beta : D(\alpha, \beta) \geq \gamma\}$, where α is a query point.

A family $\rho = \{\sigma : |S \rightarrow U\}$ is called $(\gamma_1, \gamma_2, \beta_1, \beta_2)$-sensitive for D if for any $(\alpha, \beta) \in S$

- if $\beta \in B(\alpha, \gamma_1)$, then $\Pr_\rho[\sigma(\alpha) = \sigma(\beta)] \geq \beta_1$,
- if $\beta \notin B(\alpha, \gamma_2)$, then $\Pr_\rho[\sigma(\alpha) = \sigma(\beta)] \leq \beta_2$.

In order for a locality-sensitive family to be useful, it has to satisfy inequalities $\beta_1 > \beta_2$ and $\gamma_1 < \gamma_2$, when D is a dissimilarity measure, or $\beta_1 > \beta_2$ and $\gamma_1 > \gamma_2$, when D is a similarity measure.

The working principle of LSH is shown in Fig. 1.

2.3. Comparable encryption

Comparable encryption is defined in [7], which is composed of four algorithms, Gen, Enc, Der and Cmp.

- Gen: Inputs a security parameter $\lambda \in \mathbb{N}$ and a range parameter $n \in \mathbb{N}$, outputs a parameter pa and an master key $mkey$.

  $(pa, mkey) \leftarrow Gen(\lambda, n)$

- Der: Inputs the parameter pa, the master key $mkey$, and a number $0 \leq num < 2^n$, outputs a token tok.

  $tok \leftarrow Der(pa, mkey, num)$

- Enc: Inputs pa, $mkey$, and a number $0 \leq num < 2^n$, outputs a ciphertext $ciph$.

  $ciph \leftarrow Enc(pa, mkey, num)$

- Cmp: Inputs pa, two ciphertexts $ciph$ and $ciph'$, and a token tok, outputs $\{-1, 1, 0\}$.

  $Cmp(pa, ciph, ciph', tok) \in \{-1, 1, 0\}$

We assume the ciphertext $ciph$ and the token tok input to Cmp are generated with the same parameter pa, master key $mkey$, and number num.

$tok = Der(pa, mkey, num)$

and

$ciph = Enc(pa, mkey, num)$.

The output of Cmp is $\{-1, 1, 0\}$, respectively, when

$num < num'$,

$num > num'$,

or

$num = num'$.

This requirement is formalized in the following property of completeness.

3. Retrieval model

3.1. Design goals

To enable fine-grained search for effective utilization of outsourced data in cloud, our system design should achieve performance guarantees as follows.

- **Encrypted Image Search**. To design a searchable encryption scheme which allows the clients to conduct the search of the corresponding image. And cloud server returns the corresponding results according to the required information.
- **Image Privacy**. To prevent the cloud server from learning the contents of the outsourced images.
- **Query Privacy**. To prevent the cloud server from obtaining the information of the query items.
- **Fuzzy Search**. The retrieval results are the required images according to the threshold given by the clients.

3.2. Security goals

Our system aims to achieve the following security goals.

- **Feature Vector Privacy**. The cloud server cannot recover the original feature vectors from the blind feature vectors which are generated by authorized clients.
- **Ciphertext Privacy**. The encrypted images cannot be revealed by the cloud server in the search process.
- **Correctness**. The search results are the similar images of the required feature vectors.

3.3. System model

There are three parties in the system.

- Data owner, who owns the images and can encrypt the images, generate the supplementary information of the encrypted images. Then he/she stores the ciphertexts in the remote cloud server. And the data owner shares the decryption key to the authorized clients.

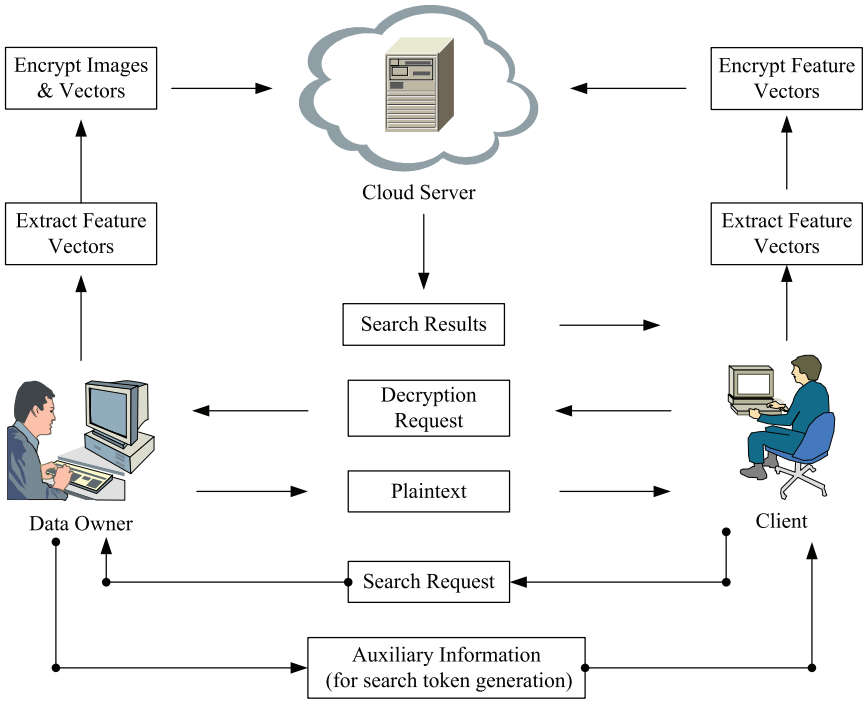

Fig. 2. Image retrieval system model.

- Cloud server, who performs the search work and returns the retrieval results to the authorized clients.
- Clients, who query for the encrypted images and obtain the decryption keys from the data owner.

We consider the following scenario.

A data owner will store some images in the remote cloud server. However, he/she do not want the cloud server know the contents of the images, thus, he/she encrypts the images before upload the images. In order to retrieval the encrypted images conveniently, the data owner upload some supplementary information with the corresponding images.

If one client wants to find some similar images of his/her own images with the feature vectors by using the matching method. In order to protect the feature vectors, the feature vectors will be encrypted. And then, he/she uploads the encrypted feature vectors to the cloud server.

The cloud server search the required feature vectors by using the retrieval algorithm, and returns the corresponding encrypted images.

After receiving the returned encrypted images, the client requests for the decryption keys of the retrieval results.

The system model is shown in Fig. 2.

4. Retrieval scheme for encrypted images

Now we introduce our retrieval algorithm for encrypted images.

- **Setup**(1^λ). A probabilistic algorithm executed by data owner to set up the system and to initialize system parameters, where λ is the security parameter. The algorithm outputs the public keys PK and secret keys SK.
- **Extract**(SK). Inputs the images, it outputs the feature vectors of each image.
- **GenIndex**(Sk, v). Inputs the images and the corresponding feature vectors v, it outputs the indices of each image.
- **Write**(qk_j, d_j). Data owner encrypts the image m_i and then invokes **GenIndex**(SK, v) to generate $I(m_i \cdot v)$. Then the data owner sends $D_i = \{E(m_i), I(m_i \cdot v)\}$ to the cloud server.
- **ConstructQ**(PK, v). Run by a client to construct a query. It takes as input the public key PK and a feature vector v, and outputs a query $Q(v)$.
- **Search**($Q(v)$, D). Run by servers to search in D associate with $Q(v)$, and outputs $R = \{E(m_i)|E(m_i) \in D, m_i \cdot v = v\}$.
- **Decrypt**(SK, $E(m_i)$). Run by a client to get the image m_i.

The framework of our scheme is shown as in Fig. 3.

Our construction is as follows:

- **Setup**. Data owner generates a finite field $\mathbb{Z}_p$, p is a prime number, and g be a generator of $\mathbb{Z}_p$. And then Data owner selects an extraction algorithm EA to extract the feature vectors of each image, and a LSH algorithm, an encryption algorithm E to encrypt the original images, and the corresponding decryption algorithm De. The public information is $\{\mathbb{Z}_p, g, \text{LSH}, \text{EA}, \text{De}\}$.
- **Extract**. Data owner computes $\text{EA}(m_i)$ of the image m_i, and gets the feature vectors $\{v_i''\}$, which are usually high dimensional vectors. Then data owner invokes the LSH algorithm to reduce the dimensions of the vectors, and gets $v_i = LSH(v_i'')$, which are the low dimensional vectors (usually the dimensions are less than 50).
- **GenIndex**. Data owner computes $l_i = |v_i|$ of each image m_i, where $|\cdot|$ is the algorithm to compute the length of the vector v. And then data owner selects a random number r_1 and blinds the feature vector v_i as

$$v_i' = r_1 \cdot v_i.$$

Another important element is l_i, which should be blind to the cloud server. The comparison encryption will be used to hide the l_i. l_i should be transformed into binary form $(b_{n-1}, b_{n-2}, \ldots, b_1, b_0)$. Data owner selects a secret key sk and generates

$$d_m = H(sk, (0, 0^\kappa, 0))$$
$$d_i = H(sk, (1, d_{i+1}, b_i))$$

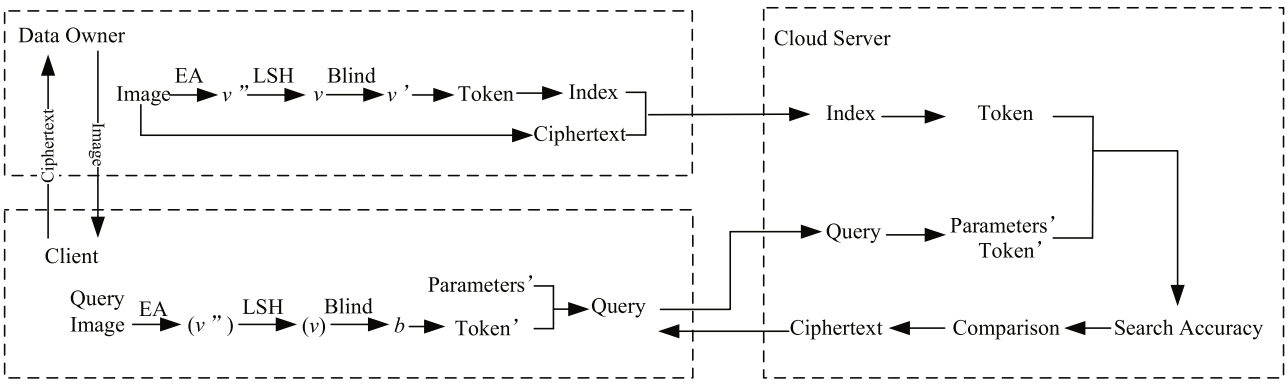

Fig. 3. Framework of our scheme.

for $i = m-1, \ldots, 0$, where κ is the length of hash result. Data owner outputs the $token_i = (d_0, d_1, \ldots, d_m)$. Then data owner randomly chooses $T \in \mathbb{Z}_p^*$, and generates

$$c_i = H(d_i, T)$$

for $i = m-1, \ldots, 0$. The index is $I(m_i) = \{T, \{v_i'\}, \{c_i\}\}$.

- **Write**. Data owner encrypts the image m_i as $E(m_i)$ and then sends $D_i = \{E(m_i), I(m_i)\}$ to the cloud server.
- **ConstructQ**. The client selects a range parameter k_1 and k_2, $k_3 \in [0, 1]$, and then invokes EA to get the feature vector v'', and gets $v' = LSH(v'')$, then blinds it as $b = r_2 \cdot v' \bmod p$, where r_2 is a random number. The client computes $l = |v'|$. The client makes a request to the data owner. If data owner agrees the search behavior, the client sends l to the data owner, and data owner generates the token of l. l should be transformed into binary form $(b'_{n-1}, b'_{n-2}, \ldots, b'_1, b'_0)$.

$$d'_m = H(sk, (0, 0^\kappa, 0))$$
$$d'_i = H(sk, (1, d'_{i+1}, b'_i))$$

for $i = m-1, \ldots, 0$.
Data owner sends the token

$$token = (d'_0, d'_1, \ldots, d'_m)$$

to the client.
The query is

$$Q(v) = \{k1, k_2, k_3, b, token\}.$$

- **Search**. When receiving the query $Q(v)$ of a client, the cloud server computes

$$c'_i = H(d'_i, T)$$

according to the index of each ciphertext D_i.
The cloud server compares the sequences c_i and c'_i, until the first different pair appears. Let $0 \le j \le m-1$, if $\forall k, j < k \le m-1$,

$$c'_k = H(d_k, T) \wedge (c'_j) \neq H(d_j, T)$$

is true, then j is the block number which two sequences are first different. This means the difference between l_i and l is no more than 2^j.
If $\forall k, 0 \le k \le m-1$,

$$c'_k = H(d_k, T)$$

holds, then $l_i = l$.
If $2^j \le k_1$, then the cloud server computes

$$\frac{v'_i \cdot b}{|v'_i||b|}.$$

If

$$\frac{v'_i \cdot b}{|v'_i||b|} \ge 1 - k_2,$$

the cloud server sets the counter $cn_i = cn_i + 1$.
For each image, the cloud server computes $\frac{cn_i}{n_i}$, where n_i is the number of the feature vectors of image m_i. If

$$\frac{cn_i}{n_i} \ge k_3,$$

the cloud server adds $E(m_i)$ to the results set R. (The search algorithm is shown in Algorithm 1.)

- **Decrypt**. When receiving the search results R, the client request the data owner for decryption.

Algorithm 1 The Search Algorithm

Input: Query $Q(v) = \{k1, k_2, k_3, b, token\}$;
Ciphertext $D_i = \{E(m_i), I(m_i)\}$;
Output: The search results.

```
for {j = 1; 1 ≤ j ≤ n_i; j++} do
  for {k = 0; 0 ≤ k ≤ m; k++} do
    c'_k = H(d'_k, T)
    if c_k ≠ c'_k; break; then
    compare 2^k with k_1
    if 2^k ≠ k_1, then
      compute v'_j·b / |v'_j||b|
      if v'_j·b / |v'_j||b| ≥ 1 − k_2 then
        cn_j = cn_j + 1;
  if cn_j / n_i ≥ k_3 then
    output E(m_i)
```

5. Security analysis

Proposition 1. *The search results satisfy the requirements of clients.*

Proof. The main retrieval principle can be shown as Fig. 4.

The length of the required vectors are determined by the equation

$$2^j \le k_1.$$

Thus, the required vectors are within the scope of the curves cv_1 and cv_2.

The angle range of the required vectors are determined by the equation

$$\frac{v'_i \cdot b}{|v'_i||b|} \ge 1 - k_2.$$

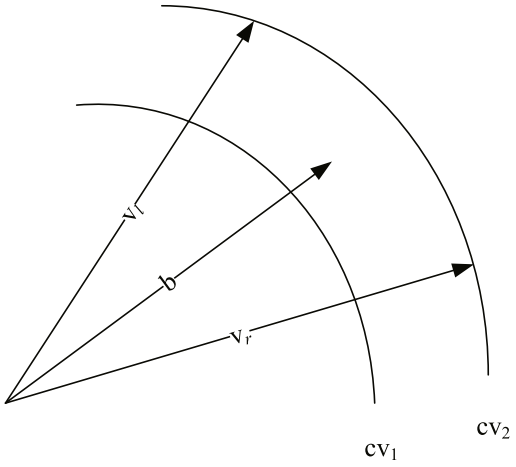

Fig. 4. Retrieval principle.

As

$$\cos\theta = \frac{v_i' \cdot b}{|v_i'||b|},$$

if the value is close to 1, the angle between the required vector and the outsourced vector is close to 0. Thus, the required vectors are within the scope of the left vector v_l and the right vector v_2.

Proposition 2. *The range* $[0, 2^j]$ *is the real range of the difference between* $|v_i|$ *and* $|v|$.

Proof. In the proposed scheme, we use comparable encryption to get the range of the difference between $|v_i|$ and $|v|$. In order to protect the value of $|v_i|$ and $|v|$, in the compare phase, they will be transformed into binary form, and the corresponding tokens will be generated. At last, the tokens are used to do the comparison.

The index of image m_i is

$$I(m_i) = \{T, \{v_i'\}, \{c_i\}\},$$

where $c_i = H(d_i, T)$, for $i = m-1, \ldots, 0$. And $\{d_i\}$ are generated as follows.

$$d_m = H(sk, (0, 0^\kappa, 0))$$
$$d_i = H(sk, (1, d_{i+1}, b_i))$$

for $i = m-1, \ldots, 0$.

In the index, v_i' is the blind vector of v_i, and $|v_i'| \neq |v_i|$. $|v_i|$ will be transformed into binary form $(b_{n-1}, b_{n-2}, \ldots, b_1, b_0)$, and in the comparison phase, only $\{c_i\}$ and T will be used.

The token of $|v|$ is generated as follows.

$|v|$ is transformed into binary form $(b'_{n-1}, b'_{n-2}, \ldots, b'_1, b'_0)$. And the compare information is

$$d'_m = H(sk, (0, 0^\kappa, 0))$$
$$d'_i = H(sk, (1, d'_{i+1}, b'_i))$$

for $i = m-1, \ldots, 0$.

In the comparison phase, the cloud server computes c'_i with the information T in $I(m_i)$ and the compare information $\{d'_i\}$.

$$c'_i = H(d'_i, T).$$

For

$$d_i = H(sk, (1, d_{i+1}, b_i)),$$

and

$$d'_i = H(sk, (1, d'_{i+1}, b'_i)).$$

Table 1
Comparisons with related works.

	Homomorphic encryption	Encrypted database
Scheme in [18]	No	No
Scheme in [27]	No	No
Scheme in [29]	Yes	Yes
Scheme in [17]	No	No
Our Scheme	No	Yes

Table 2
Computation cost.

	GenIndex	GenQuery	Search
MM	y	0	$3y$
MInv	0	0	y
Hash	$2my$	m	my

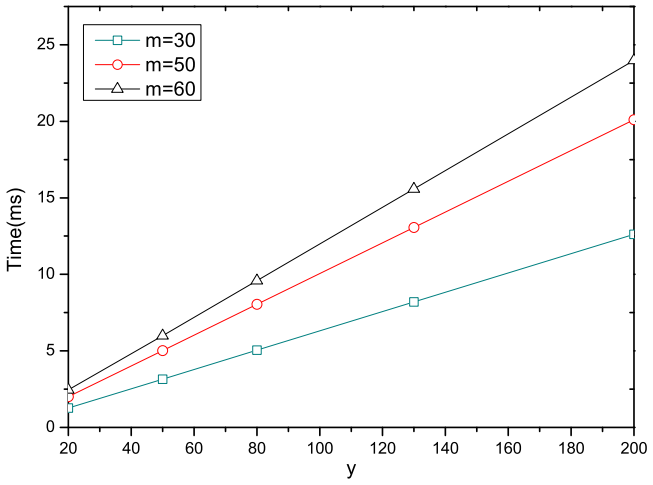

Fig. 5. Time cost in GenIndex phase.

If $b_i = b'_i$ and $d_{i+1} = d'_{i+1}$, then $d_i = d'_i$, thus, $c_i = c'_i$.

For $0 \leq j \leq m-1$, if $\forall k, j < k \leq m-1$,

$$c'_k = H(d_k, T) \wedge (c'_j) \neq H(d_j, T)$$

is true, then $c_{j+1} = c'_{j+1}$ and $c_j \neq c'_j$.

This means $d_s = d'_s$, and $d_j \neq d'_j$, for $j \leq s \leq m-1$.

Then $b_s = b'_s$, and $b_j \neq b'_j$, for $j \leq s \leq m-1$.

For b_j and b'_j are binary values, the difference between b_j and b'_j is 1.

Thus, the difference between $|v'_i|$ and $|v|$ is no more than 2^j.

6. Comparisons and efficiency

Features comparisons between our scheme and some recent schemes are list in Table 1.

Let MM denote modular multiplication, MInv denote modular inverse. Let m be the maximum length of the modular of feature vectors in binary form, and y be the number of encrypted images. For every feature vector, the computation cost in index generation Phase, query generation phase and search phase are shown in Table 2.

We implement our mechanism using C language and pairing-based cryptography(PBC)library. The process is conducted on a computer with Intel Xeon E5-1620 CPU processor running at 3.50 GHz, 16 GB RAM. The time cost in GenIndex phase, GenQuery phase and Search phase are shown in Figs. 5–7.

Though the computation cost in index generation phase seems a little more, it is the initialization work which is done for only once.

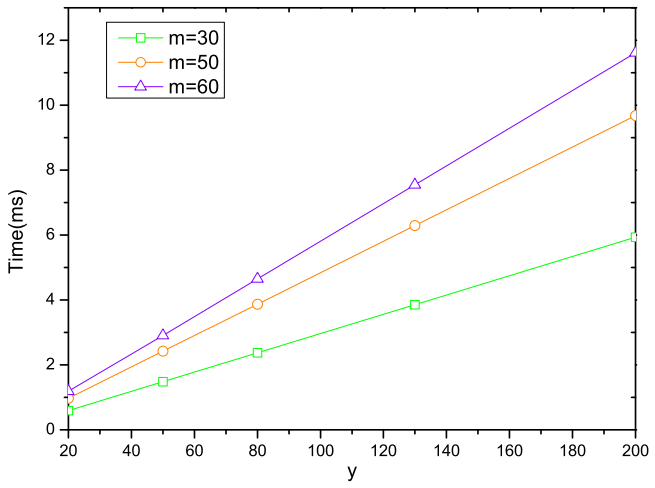

Fig. 6. Time cost in GenQuery phase.

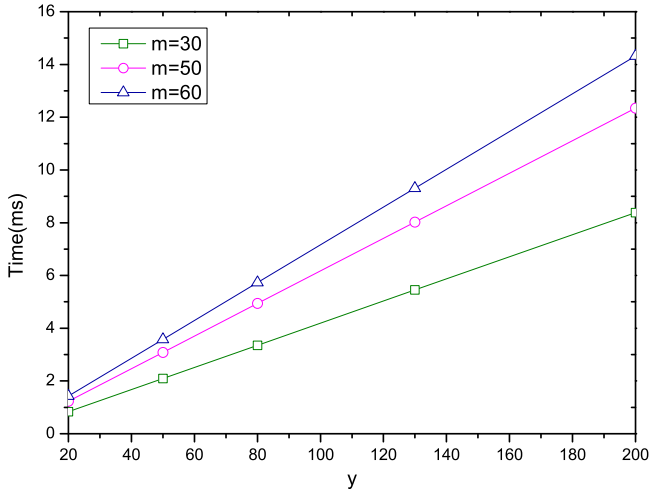

Fig. 7. Time cost in Search phase.

However, in the query generation phase (the repetitive work), the computational cost is small. The heavy work in search phase is left for the powerful cloud server, which can implement the search work easily.

7. Conclusion

With the development of cloud computing, the outsourced storage is more and more popular. The powerful cloud server provides huge storage space for clients. In order to protect the privacy of the outsourced information, the client has to encrypt the information, and uploads the ciphertexts. However, it is hard to search the required information in the encrypted files. In this paper, an efficient retrieval scheme for encrypted images is proposed, and a novel search model is introduced. The retrieval scheme is flexible. The users who get the authorization by data owner can search the encrypted images. In the proposed scheme, the feature vectors are blind to the cloud server, and the length of the feature vector is keep private. On the other hand, in order to improve the search efficiency, and reduce the computational cost of clients, the homomorphic encryption is not used in our scheme. The communication cost is low, only one-round interaction is needed to search for target images between the cloud and the clients. Our scheme can be used in encrypted database, the private information of the images can be kept secure.

Acknowledgments

This work was supported by the Fund of Lab of Security Insurance of Cyberspace, Sichuan Province (szjj2016-091); Guangxi Key Laboratory of Cryptography and Information Security (No. GCIS201719); the Talent Project of Sichuan University of Science & Engineering (2017RCL23); the Science Founding of Artificial Intelligence Key Laboratory of Sichuan Province (2017RZJ03); the Opening Project of Sichuan Province University Key Laboratory of Bridge Non-destruction Detecting and Engineering Computing (2017QYJ03); the CCF funding (No. CCF-VenustechRP2017006); the key laboratory of higher education of Sichuan Province for enterprise informationalization and Internet of things (No. 2017WYJ02).

References

[1] Y. Teing, A. Dehghantanha, K.K.R. Choo, L.T. Yang, Forensic investigation of P2P cloud storage services and backbone for iot networks: Bittorrent sync as a case study, Comput. Electrical Eng. 58 (2017) 350–363.
[2] N.H.A. Rahman, N.D.W. Cahyani, K.K.R. Choo, Cloud incident handling and forensic-by-design: cloud storage as a case study, Concurr. Comput.: Pract. Exp. 29 (14) (2017).
[3] B. Martini, K.K.R. Choo, Cloud forensic technical challenges and solutions: A snapshot, IEEE Cloud Comput. 1 (14) (2014) 20–25.
[4] B. Martini, K.K.R. Choo, Distributed filesystem forensics: Xtreemfs as a case study, Digital Investigation 11 (4) (2014) 295–313.
[5] Y. Hu, K.K.R. Choo, W. Chen, Tamper detection and image recovery for BTC-compressed images, Multimedia Tools Appl. 76 (14) (2017) 15435–15463.
[6] S.A. Miraftabzadeh, P. Rad, K.K.R. Choo, M. Jamshidi, A privacy-aware architecture at the edge for autonomous real-time identity re-identification in crowds, IEEE IoT J. (2018). http://dx.doi.org/10.1109/JIOT.2017.2761801.
[7] J. Furukawa, Request-Based comparable encryption, in: Computer Security - ESORICS 2013 - 18th European Symposium on Research in Computer Security, Egham, UK, September 9–13, 2013. Proceedings, vol. 8134, 2013, pp. 129–146.
[8] C. Wang, S. Chow, Q. Wang, K. Ren, W. Lou, Privacy-preserving public auditing for secure cloud storage, IEEE Trans. Comput. 62 (2) (2013) 362–375.
[9] J. Wu, L. Ping, X. Ge, Y. Wang, J. Fu, Cloud storage as the infrastructure of cloud computing, in: International Conference on Intelligent Computing and Cognitive Informatics, ICICCI, Kuala Lumpur, Malaysia, June 22–23, 2010, pp. 380–383.
[10] G. Ateniese, Ö. Dagdelen, I. Damgård, D. Venturi, Entangled cloud storage, Future Gener. Comput. Syst. 62 (2016) 104–118.
[11] S.D.X., W.D., P.A., Practical techniques for searches on encrypted data, in: 2000 IEEE Symposium on Security and Privacy, Berkeley, California, USA, May 14–17, 2000, Springer-Verlag, 2000, pp. 44–55.
[12] Y. Chang, M. Mitzenmacher, Privacy preserving keyword searches on remote encrypted data, in: Applied Cryptography and Network Security, Springer, 2005, pp. 442–455.
[13] C.R., G.J.A., K.S., O.R., Searchable symmetric encryption: improved definitions and efficient constructions, in: Proc. of the 13th ACM Conference on Computer and Communications Security, CCS 2006, Alexandria, VA, USA, Ioctober 30 - November 3, 2006, pp. 79–88.
[14] K. Yee, K. Swearingen, K. Li, M.A. Hearst, Faceted metadata for image search and browsing, in: Proceedings of the 2003 Conference on Human Factors in Computing Systems, CHI 2003, Ft. Lauderdale, Florida, USA, April 5–10, 2003, pp. 401–408.
[15] Y. Jing, S. Baluja, Visualrank: Applying pagerank to large-scale image search, IEEE Trans. Pattern Anal. Mach. Intell. 30 (11) (2008) 1877–1890.
[16] Y. Kiyoki, T. Kitagawa, T. Hayama, A metadatabase system for semantic image search by a mathematical model of meaning, SIGMOD Record 23 (4) (1994) 34–41.
[17] G. Liu, J. Yang, Content-based image retrieval using color difference histogram, Pattern Recognit. 46 (1) (2013) 188–198.
[18] J. Guo, H. Prasetyo, J. Chen, Content-based image retrieval using error diffusion block truncation coding features, IEEE Trans. Circuits Syst. Video Technol. 25 (3) (2015) 466–481.
[19] P. Enser, Progress in documentation pictorial information retrieval, J. Doc. 51 (2) (1995) 126–170.
[20] G. Salton, M. McGill, Introduction to Modern Information Retrieval, McGraw-Hill Book Company, 1984.

[21] E. Chatzistavros, S.A. Chatzichristofis, K. Zagoris, G. Stamatelos, Content-based image retrieval over IEEE 802.11b noisy wireless networks, Int. J. Commun. Syst. 28 (8) (2015) 1432–1449.
[22] R. Datta, J. Li, J. Wang, Content-based image retrieval: approaches and trends of the new age, in: Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval, ACM, 2005, pp. 253–262.
[23] E. de Ves, X. Benavent, I. Coma, G. Ayala, A novel dynamic multi-model relevance feedback procedure for content-based image retrieval, Neurocomputing 208 (2016) 99–107.
[24] L. Al-Safadi, R. Alomran, F. Almutairi, An overview and evaluation of the radiologists lounge, a semantic content-based radiographic images retrieval, Multimedia Tools Appl. 75 (1) (2016) 607–625.
[25] W. Lu, A. Swaminathan, A. Varna, M. Wu, Enabling search over encrypted multimedia databases, in: Media Forensics and Security I, Part of the IS&T-SPIE Electronic Imaging Symposium, San Jose, CA, USA, January 19–21, 2009, Proceedings, 2009, p. 725418.
[26] L. Weng, L. Amsaleg, A. Morton, S. Marchand-Maillet, A privacy-preserving framework for large-scale content-based information retrieval, IEEE Trans. Inf. Forensics Secur. 10 (1) (2015) 152–167.
[27] Y. Chun, N. Kim, I. Jang, Content-based image retrieval using multiresolution color and texture features, IEEE Trans. Multimedia 10 (6) (2008) 1073–1084.
[28] Y. Chen, X. Li, A. Dick, R. Hill, Ranking consistency for image matching and object retrieval, Pattern Recognit. 47 (3) (2014) 1349–1360.
[29] R. Bellafqira, G. Coatrieux, D. Bouslimi, G. Quellec, Content-based image retrieval in homomorphic encryption domain, in: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2015, Milan, Italy, August 25–29, 2015, 2015, pp. 2944–2947.
[30] W. Lu, A.L. Varna, A. Swaminathan, M. Wu, Secure image retrieval through feature protection, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009, 19–24 April 2009, Taipei, Taiwan, 2009, pp. 1533–1536.
[31] B. Cheng, L. Zhuo, Y. Bai, Y. Peng, J. Zhang, Secure index construction for privacy-preserving large-scale image retrieval, in: 2014 IEEE Fourth International Conference on Big Data and Cloud Computing, BDCloud 2014, Sydney, Australia, December 3–5, 2014, pp. 116–120.

Jun Ye received his B.S. degree in Applied Mathematics at Chongqing University. M.S. degree in Cryptography at Guilin University of Electronic Technology. He is a Lecturer at the School of Science, Sichuan University of Science & Engineering. His current research interests include cryptography and information security.

Zheng Xu was born in Shanghai, China. He received the Diploma and Ph.D. degrees from the School of Computing Engineering and Science, Shanghai University, Shanghai, in 2007 and 2012, respectively. He is currently working in the third research institute of ministry of public security and the postdoctoral in Tsinghua University, China. His current research interests include topic detection and tracking, semantic Web and Web mining. He has authored or co-authored more than 70 publications including IEEE Trans. On Fuzzy Systems, IEEE Trans. On Automation Science and Engineering, IEEE Trans. On Cloud Computing, IEEE Trans. On Emerging Topics in Computing, IEEE Transactions on Systems, Man, and Cybernetics: Systems, etc.

Yong Ding was born in Chongqing, China. He received the Ph.D. degrees from the School of Communication Engineering, Xidian University, Shaanxi, in 2005. He is currently professor in Guilin University of Electronic Technology, China. His current research interests include cryptography and information security.