Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Content-based multi-source encrypted image retrieval in clouds with privacy preservation

Meng Shen [a,*], Guohua Cheng [a], Liehuang Zhu [a], Xiaojiang Du [b], Jiankun Hu [c]

[a] School of Computer Science, Beijing Institute of Technology, Beijing 100081, China
[b] Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122, USA
[c] School of Engineering and IT, University of New South Wales (UNSW), Canberra, ACT 2610, Australia

HIGHLIGHTS

- The scheme enables content-based retrieval from multi-source encrypted images.
- Multiple image owners are allowed to encrypt images by their unique secret keys.
- A new approach to measuring the similarity of encrypted images is proposed.

ARTICLE INFO

Article history:
Received 29 September 2017
Received in revised form 9 March 2018
Accepted 25 April 2018
Available online xxxx

Keywords:
Secure image retrieval
Multi-source
Privacy preserving
Searchable encryption
Content-based image retrieval
Image encryption

ABSTRACT

Content-based image retrieval (CBIR) is one of the fundamental image retrieval primitives. Its applications can be found in various areas, such as art collections and medical diagnoses. With an increasing prevalence of cloud computing paradigm, image owners desire to outsource their images to cloud servers. In order to deal with the risk of privacy leakage of images, images are typically encrypted before they are outsourced to the cloud, which makes CBIR an extremely challenging task. Existing studies focus on the scenario with only a single image owner, leaving the problem of CBIR with multiple image sources (i.e., owners) unaddressed.

In this paper, we propose a secure CBIR scheme that supports Multiple Image owners with Privacy Protection (MIPP). We encrypt image features with a secure multi-party computation technique, which allows image owners to encrypt image features with their own keys. This enables efficient image retrieval over images gathered from multiple sources, while guaranteeing that image privacy of an individual image owner will not be leaked to other image owners. We also propose a new method for similarity measurement of images that can avoid revealing image similarity information to the cloud. Theoretical analysis and experimental results demonstrate that MIPP achieves retrieval accuracy and efficiency simultaneously, while preserving image privacy.

1. Introduction

Recent years have witnessed the prosperity of image-sharing services and applications (e.g., Instagram), which results in an increasing demand for image retrieval. In early years, text-based image retrieval systems implemented by manual tagging image properties could fulfill the requirement of image retrieval. With the growing popularity of Internet users, hundreds of millions of images appear on the Internet per second, the traditional text-based image retrieval becomes gradually impractical, because it consumes prohibitive manpower and financial resources for labeling. Content-based image retrieval (CBIR) [1–3] has been proposed for real world applications which uses the image feature extracted automatically from images, such as colors [4,5], textures [6,7], and shapes [8,9].

In general, higher resolution images consume more storage. For instance, a photo taken by a cell phone in present days may be about 2 MB and the one taken by a professional camera may reach 10 MB or more. With an increasing prevalence of cloud computing and storage [10,11], migrating services to the cloud has rapidly become a trend for mass data storage and management. By outsourcing images to the cloud, service providers can make their services easily accessible to geographically distributed users by only requiring them to pay for the computation and storage resources they actually use.

* Corresponding author.
E-mail addresses: shenmeng@bit.edu.cn (M. Shen), chengzi92036@163.com (G. Cheng), liehuangz@bit.edu.cn (L. Zhu), dxj@ieee.org (X. Du), J.Hu@adfa.edu.au (J. Hu).

https://doi.org/10.1016/j.future.2018.04.089
0167-739X/

Outsourcing images directly to cloud servers, however, increases the risk of privacy leakage when images contain sensitive information, such as patient's medical information or personal location information. For instance, a compromised cloud vendor could enable access to the outsourced images by unauthorized users. In order to protect images against privacy leakage threats, images are usually encrypted before being outsourced to the cloud. Since encryption operations disrupt the image content, it becomes a challenging task to perform CBIR over encrypted images. Therefore, it is highly desirable to devise a privacy-preserving CBIR system for cloud-based encrypted image sets.

Many schemes have been proposed in the field of secure CBIR [12–22], which can be roughly classified into two categories. In the first category, image owners extract features from plain images, and then outsource both the encrypted images and the encrypted image features to the cloud. In the second category, image owners outsource only the encrypted images to the cloud that is responsible for extracting features from encrypted images and for conducting retrieval operations.

Existing studies have a common limitation that they consider only a single-source case (i.e., a single image owner). In real-world applications, however, image retrieval is more likely to get multiple image sources involved. For instance, consider a cloud-based e-health application, which takes an encrypted ultrasonic medical image of an undiagnosed patient as input and searches for similar confirmed cases from a collection of encrypted medical images. Suppose images are collected from multiple hospitals (i.e., sources), each of which is reluctant and unpermitted to share with one another the plain medical images. The existing schemes can be easily extended to the multi-source scenario by performing retrieval over encrypted images of different owners *one by one*. Although simple and straightforward, it introduces multiple rounds of communications between users and individual image owners, and thereby becomes inefficient in terms of retrieval time and communication overhead.

There are several challenges in designing a secure and efficient CBIR scheme with multiple image owners. First, we should ensure the privacy of images and image features of different image owners. Second, the authorized query user should communicate his secret image encryption key with image owners for generating a secret query in secure image retrieval schemes. However, when secure image retrieval schemes have multiple image owners, each image owner should use their own secret image encryption key to encrypt images and image features. Then, the authorized query user should communicate his secret image encryption key with each image owner for generating a secret query, which will increase the communication overload in schemes. It is desirable to address this problem in the secure image retrieval scheme with multiple image owners. Finally, when the cloud executes image retrieval, it may obtain similarity relation information of images in the retrieval result. This privacy issue should be also addressed.

In this paper, we propose the MIPP, a novel content-based multi-source image retrieval scheme with privacy protection. MIPP operates in the same way as existing schemes in the first category, which outsources encrypted images along with their encrypted image features to the cloud. In order to address the challenges of supporting multiple image owners, we first encrypt images with a key stream and encrypt the corresponding image features by the secure multi-party computation method, and then propose a novel method to measure the image similarity; this can help to avoid revealing the image similarity information in cloud to a certain extent.

The main contributions in this paper are highlighted as follows:

- We design a MIPP, which, to the best of our knowledge, is the first scheme belonging to the first category that enables content-based multi-source image retrieval with privacy protection. In the proposed MIPP, multiple image owners are allowed to encrypt images and image features by their unique secret image encryption keys. This enables an efficient image retrieval over images gathered from multiple sources, while providing guarantees that image privacy of an individual image owner will not be leaked to other image owners. Thus, the proposed MIPP can meet the practical requirements in real-world applications.
- We present a new approach to measure the similarity of images, which can avoid the leakage of image similarity information in retrieval results. Extensive experimental results show that the resulting retrieval outcome is comparable to that with the typical Euclidean distance criterion.

The rest of this paper is organized as follows. We summarize the related work in Section 2 and introduce the preliminaries in Section 3. In Section 4, we present the system model, thread model, and design goals of our scheme. We detail the design of MIPP in Section 5 and present the security analysis in Section 6. We evaluate the performance of the proposed scheme in Section 7 and conclude this paper in Section 8.

2. Related work

In this section, we will present a brief overview of existing research schemes in the field of secure image retrieval.

Homomorphic encryption is a form of encryption that allows computations to be carried out on ciphertext, thus generating an encrypted result which, when decrypted, matches the result of operations performed on the plaintext. Zhang et al. [12,13] leveraged the property of homomorphic encryption in secure image retrieval. The homomorphic encryption results in high computational complexity that makes it consume too much time. Xia et al. [14] proposed a secure CBIR scheme with Bag-of-Words model and Earth Movers Distance. The retrieval index was constructed by locality-sensitive hashing. In this scheme, the user and image owner have two times of two-way communications, which results in high communication overhead. Yuan et al. [15] proposed a scheme named SEISA with access control and secure k-means outsource, dynamically updating images is supported. Yuan et al. [16] proposed a scheme that can explore user relationship while preserving image privacy. The secure index and encrypted image features were constructed by an entity called SF rather the image owner. This scheme supports dynamic updates of images without affecting the current social structure. The scheme proposed by Xia et al. [17] was able to deter the illegal distribution of images while preserving the image privacy. Li et al. [18] proposed a privacy preserving retrieval scheme for outsourced media, which used the one-way hash along with encrypting partly hash values to encrypt image features. This scheme created trade offs among privacy preserving, retrieval quality, and complexity through adjusting the bit counts of encryption in the hash value.

In the second category, Ferreira et al. [19] proposed a scheme IES-CBIR that can extract image features from encrypted images. The texture and color features were encrypted separately; the color feature was encrypted by scrambling pixels in HSV color model and the texture feature is encrypted by shuffling rows and columns in images. This scheme enabled dynamic updates of images by using Bag-Of-Visual-Words model. Zhang et al. [20] proposed a histogram-based retrieval scheme for encrypted JPEG images with machine learning. They encrypted images by permuting DCT and the server to retrieve the histogram at each frequency position from encrypted images. Cheng et al. [21] proposed a Markov process based retrieval scheme for encrypted JPEG images. Markov's process models of the AC coefficients and the server could extract

Please cite this article in press as: M. Shen, et al., Content-based multi-source encrypted image retrieval in clouds with privacy preservation, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.04.089.

features from the transition probability matrices of those AC coefficients of encrypted images. Zhang et al. [22] proposed an encrypted medical image retrieval algorithm based on DWT–DCT frequency domain. In this algorithm, features were extracted from encrypted images.

As described in Section 1, these schemes may lead to heavy communication overload and privacy leakage (e.g., image features and image similarity) when simply extended for CBIR with multiple image owners.

3. Preliminary

In this section, we will introduce the preliminaries, including image features used in this paper and the secure multi-party computation.

3.1. Image feature

MPEG-7 [23,24] standard is the multi-media content description interface that contains a set of descriptors. We extract the Edge Histogram Descriptor (EHD) from images for our secure CBIR scheme. Edge Histogram Description is a non-homogeneous texture descriptor in MPEG-7 which captures spatial distribution of edges and works well in CBIR.

3.2. Secure multi-party computation

Assume that there are multiple parties, each of which owns a secret number. Each party expects to obtain the total number of all parties, without publicizing their own number to others. Thus, each participant needs to encrypt his number before making it public. The secure multi-party computation can calculate over encrypted numbers, which meets the above requirement. Under the background of the secure multi-party computation technology, we can obtain the sum or product of numbers in the same manner as calculating under plain number. Nowadays, secure multi-party computation has been applied in many real-world applications, such as secure voting and secure electronic auction.

Jung et al. [25] proposed a privacy-preserving sum calculation scheme with collusion-tolerable, without the need for secure channels. This scheme can effectively calculate the sum of encrypted numbers, which can be briefly described as follows.

Step 1: Select two large prime numbers p and q with the same length where q divides $p-1$.

Step 2: Define the q-order cyclic multiplicative group G_1 with a generator being defined in Eq. (1), where h is a random number in Z_p. And define the q-order multiplicative group G_2, where its generator is defined in Eq. (2).

$$g_1 = h^{(p-1)/q} \mod p \quad s.t. \quad g_1 \neq 1 \mod p \tag{1}$$

$$g_2 = g_1^p \mod p^2 \tag{2}$$

Step 3: Each participant P_i randomly chooses a number $r_i \in Z_q$ and calculates a public number ${g_2}^{r_i} mod p^2$. Then, she exchanges the number ${g_2}^{r_i} \in G_2$ with P_{i-1} and P_{i+1}. P_i can calculate a secret number $R_i \in G_2$ as shown in Eq. (3) and the ciphertext C_i as shown in Eq. (4) after a round of exchanges.

$$R_i = (g_2^{r_{i+1}} / g_2^{r_{i-1}})^{r_i} \tag{3}$$

$$C_i = (1 + x_i p) R_i \mod p^2 \tag{4}$$

Step 4: Each P_i shares her ciphertext C_i with other participants and calculates the product of all ciphertexts according to Eq. (5) to obtain the value of C.

$$\begin{aligned} C &= \prod_{i=1}^{n} C_i \mod p^2 \\ &= \prod_{i=1}^{n} (1 + x_i p)(g_2^{r_{i+1}} / g_2^{r_{i-1}})^{r_i} \mod p^2 \\ &= (1 + p \sum_{i=1}^{n} x_i) g_2^{\sum_{i=1}^{n} r_{i+1} r_i - r_i r_{i-1}} \mod p^2 \\ &= (1 + p \sum_{i=1}^{n} x_i) \mod p^2 \end{aligned} \tag{5}$$

Step 5: The sum of all numbers can be obtained by calculating Eq. (6).

$$(C-1)/p = \sum_{i=1}^{n} x_i \mod p \tag{6}$$

4. Problem formulation

In this section, we will introduce the system model, threat model, and design goals of our scheme.

4.1. System model

There are four types of entities in our secure multi-source CBIR system, including multiple image owners, authorized query users, the cloud, and a key management center (KMC), as illustrated in Fig. 1. The description of each type of entity is detailed as follows.

- **Multiple images owners**: They are providers of image databases denoted by Image Owner i ($i \in [1, N]$) in Fig. 1. We assume that each image owner has a secure channel to communicate his secret key with KMC.
- **Authorized query users**: They are users authorized by specific image owners and have the authority to send image retrieval requests to the cloud. We assume that authorized query users in the system will not reveal their secret image encryption key or distribute image retrieval results to unauthorized users.
- **Cloud**: It takes responsibility for building secure retrieval indexes and executing image retrieval. It stores encrypted images, encrypted image features, information of image owners, and a list of authorized query users. We assume that the cloud is honest-but-curious, which means it will execute the image retrieval operation correctly, while it may also analyze images or image features to obtain some sensitive information about images.
- **Key management center (KMC)**: It has three main functionalities. First, it takes responsibility for storing secret image encryption keys, information of image owners, authorized query user lists received from image owners, and storing the authorized query user's secret image encryption key temporarily when there comes a query from an authorized query user. Second, it will decrypt encrypted image retrieval results from the cloud, and then encrypt these images with the secret image encryption key of the authorized query user. Finally, it will send the new encrypted image retrieval results to the cloud. In this paper, we assume that the KMC is fully trusted that it will not reveal secret image encryption keys of image owners and authorized query users to others.

Please cite this article in press as: M. Shen, et al., Content-based multi-source encrypted image retrieval in clouds with privacy preservation, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.04.089.

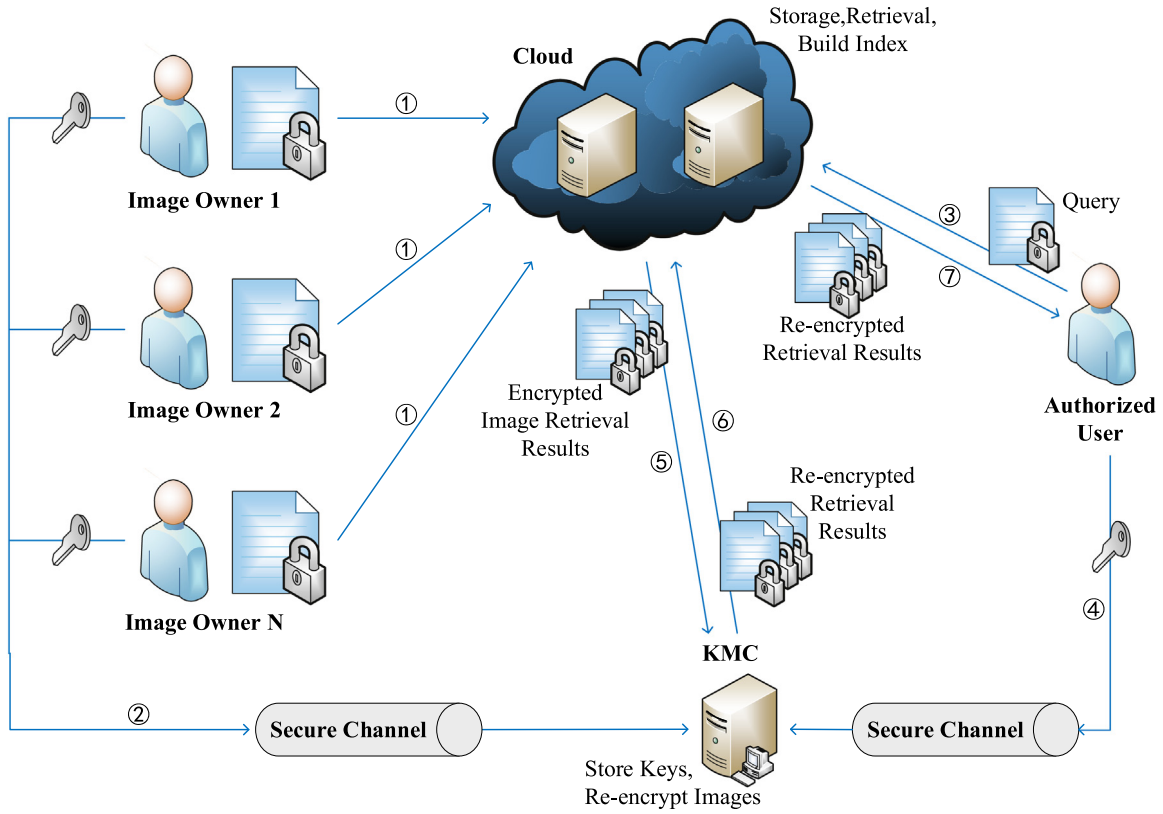


Fig. 1. System Model of the Secure Multi-Source CBIR Scheme.

The workflow of our scheme is described as follows:

(1) Multiple image owners extract the EHD feature to represent an image, then they encrypt their images and image features respectively. After that, they outsource encrypted images and encrypted image features along with their identity to the cloud. The authorized user list will also be sent to the cloud for the image retrieval service, shown as ① in Fig. 1. They also need to send secret image encryption keys that is used to encrypt images to the KMC through a secure channel, shown as ② in Fig. 1. Image owners need not store secret keys that is used to encrypt image features. When a new image owner arrives, the above operations should be repeated.

(2) Before an authorized user requests a query, he/she will extract EHD features from query images and encrypt query image features, then submit the generated encrypted query to the cloud for image retrieval operation, shown as ③ in Fig. 1. At the same time, he should send his secret image encryption key to the KMC through a secure channel, shown as ④ in Fig. 1. For each retrieval request, the image encryption key that the authorized query user sends to the KMC should be different from the last time he retrieved. After receiving retrieval results from cloud, the authorized query user decrypts retrieval results with his own image encryption key, which is the same as the key sent to the KMC. Finally, the authorized query user can sort the decrypted retrieval results to get top-h similar images.

(3) After receiving encrypted images and image features from image owners, the cloud will build the retrieval index. When an image retrieval request arrives, he should verify the identity of query user. Then, it will execute an image retrieval operation if verified successfully. When it obtains retrieval results, it will send the encrypted retrieval results to the KMC instead of sending the retrieval results to the authorized query user, shown as ⑤ in Fig. 1. At the same time, information of the authorized query user will also be sent to the KMC along with the corresponding encrypted retrieval results.

(4) When the KMC receives encrypted images from cloud, it will decrypt these images with image owners' secret image encryption key first. Then, it will encrypt these images with the secret image encryption key of the related authorized query user. After that, the re-encrypted images will be sent to the cloud, shown as ⑥ in Fig. 1.

(5) Cloud will return the re-encrypted image retrieval results that is received from the KMC to the authorized query user, shown as ⑦ in Fig. 1.

4.2. Threat model

In our scheme, we consider the following two kinds of threads:

(1) Eavesdroppers.
In the process of image transmission (e.g., sending encrypted images and their features to the cloud, and fetching retrieval results from the KMC, the KMC sending re-encrypted images to the cloud, and the cloud providing retrieval results to the authorized query user), eavesdroppers may eavesdrop image information. This is a weak adversary that can be defended by encryption.

(2) Cloud.

Please cite this article in press as: M. Shen, et al., Content-based multi-source encrypted image retrieval in clouds with privacy preservation, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.04.089.

Table 1
Notations used in our scheme.

Notation	Description
n	The size of images and image features
u	The size of query image and image features.
OID	The identity of each image owner
AUL	The authorized user list
SK	The secret image encryption key of each image owner
USK	The secret image encryption key of query user
UID	The user id
AK	The authentication key to verify query user's identity
$W = \{w_1, w_2, \ldots, w_n\}$	The plain image collection of each image owner
$F = \{f_1, f_2, \ldots, f_n\}$	The plain image feature collection of each image owner
$FF = \{f_1^2, f_2^2, \ldots, f_n^2\}$	The square image feature collection of each image owner
$EW = \{ew_1, ew_2, \ldots, ew_n\}$	The encrypted image collection of each image owner
$EF = \{ef_1, ef_2, \ldots, ef_n\}$	The encrypted image feature collection of each image owner
$EFF = \{ef_1^2, ef_2^2, \ldots, ef_n^2\}$	The encrypted square image feature collection of each image owner
$QW = \{qw_1, qw_2, \ldots, qw_u\}$	The query image collection of each query user
$QF = \{qf_1, qf_2, \ldots, qf_u\}$	The query image feature collection of each query user
$EQ = \{eq_1, eq_2, \ldots, eq_u\}$	The encrypted query image feature collection of each query user
$QWW = \{qf_1^2, qf_2^2, \ldots, qf_u^2\}$	The square query image feature collection of each query user
$EQQ = \{eq_1^2, eq_2^2, \ldots, eq_u^2\}$	The encrypted square query image feature collection of each query user
$S = \{s_1, s_2, \ldots, s_h\}$.	The top-h retrieval results
$ER = \{er_1, er_2, \ldots, er_h\}$	The top-h encrypted retrieval results
$NER = \{ner_1, ner_2, \ldots, ner_h\}$	The re-encrypted retrieval results

In our scheme, we assume the cloud is honest-but-curious. It will correctly execute the image retrieval operation, but it may analyze image content through encrypted images with their features at the same time. Thus, we should guarantee the data privacy in the process of encrypting images and image features. Additionally, the cloud may obtain the similarity relation information of images over the process of secure image retrieval. We should avoid this kind of image privacy leakage in cloud.

4.3. Design goals

In this section, we describe design goals of our scheme as follows.

(1) Image privacy.
Image privacy is very important in secure image retrieval service. We should unable the cloud and unauthorized users to obtain plain images and plain image features, along with image similarity relation information through encrypted images, encrypted image features, the encrypted retrieval results.
(2) Retrieval accuracy.
The retrieval accuracy is an indispensable element in secure image retrieval. In our scheme, we proposed a new approach to manage the similarity of images. Therefore, retrieval accuracy difference between our scheme and the secure scheme retrieval with Euclidean distance should be within reasonable limits.
(3) Efficiency.
Efficiency indicates the time consumption in a secure image retrieval scheme. We should ensure high efficiency in our scheme to enhance its practicality in real world application, which means the time of encryption, index construction, and retrieval should be reduced.

5. The design of MIPP

In this section, notations in our scheme are shown in the table. We first introduce the system overview and then describe the data encryption method. In order to solve the image similarity leakage problem in the cloud, we propose a new method to measure image similarity. The secure content-based image retrieval, index construction, and index update method of our scheme are also introduced in this section.

5.1. Notations in this section

Notations used in our scheme are described in Table 1.

5.2. System overview

There are four entities in our system, each with its own responsibility, described as follows.

- **Image owner** is the provider of image database. Each image owner has his own image collection W. He will extract EHD image features from W and get the image feature collection F and FF. For preserving the privacy of images and image features, he will generate a secret image encryption key SK and run the **ImageEnc** algorithm to get encrypted image collection EW first. Then, he should generate secret image feature encryption keys and run the **ImageFeatureEnc** process to get the encrypted image feature collection EF and encrypted collection EFF. Then EW, EF and EFF will be outsourced to the cloud along with OID, AUL and AK through the network. Besides, SK will also be sent to the KMC for storage through a secure channel.
- **The authorized user** is the user who has the demand of image retrieval. He will extract EHD features from query images QW to obtain the query image feature collection QF and run **ImageFeatureEnc** process to get the encrypted image feature collection EQ firstly. Next, he will make calculations to obtain the collection QWW and EQQ. After that, he will generate a secret query $Q = \{EQ, EQQ, UID, AK\}$ and send Q to the cloud for secure image retrieval. Finally, he will generate a secret image encryption key USK and send this key to KMC through a secret channel. After receiving retrieval results from cloud, he uses his own secret image encryption key USK and runs the **ImageDec** algorithm to decrypt the result images ER, then sort the result images to obtain top-h similar images $S = \{s_1, s_2, \ldots, s_h\}$.

Please cite this article in press as: M. Shen, et al., Content-based multi-source encrypted image retrieval in clouds with privacy preservation, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.04.089.

- **The cloud** takes responsibility for storing and retrieving. It will store *EW*, *EF*, *OID*, *AUL* and *AK*. Given a query, it uses *AK* to verify query user's identity in the AUL. If validated successfully, it will run the **ImageRetrieval** process to retrieve similar images in the image database. Instead of sending top-h encrypted retrieval results *ER* to authorized query user directly, he sends *ER* and *AK* to KMC firstly. After that, he will send *NER* that is received from *KMC* to the authorized query user. Last, for improving retrieval efficiency, it will run **IndexConstruct** process to construct the retrieval index *I*.
- **The KMC** stores *SK*, *AUL* and temporarily stores *USK*. After receiving *ER* and *AK* from cloud, it will run the **ImageDec** algorithm to decrypt *ER* to obtain *W*, then it uses the *AK* to obtain the secret image encryption key of the authorized query user and runs the **ImageEnc** algorithm to encrypt *W* with this key to obtain *NER*. Finally, it will send *NER* to the cloud. After it finish the above operations, it can discard the *USK* of this query.

5.3. Data encryption

For preserving image privacy in the cloud, images and image features should be encrypted before outsourcing to the cloud. We will introduce data encryption methods of our scheme in this section, including key generation, image encryption, image decryption, and image feature encryption method.

5.3.1. Key generation

Given a secret parameter *k*, run the *KeyGen*(1^k) algorithm, so image owners and authorized query users can obtain the secret image encryption key *SK* and *USK* respectively, where the length of *SK* and *USK* is at least the same as the total number pixels in images and the numbers in *SK* and *USK* is between 0 and 255 inclusively.

5.3.2. ImageEnc&ImageDec — image encryption and decryption

Given a secret key *SK* and an image collection *W*, image owners run the *ImageEnc(SK, W)* algorithm to encrypt images, shown as Algorithm 1. The authorized query user also runs this algorithm to encrypt images with his secret key *USK*. In our image encryption scheme, we use a standard key stream to encrypt images which is secure against the Chosen-Plaintext Attacks (CPA). Thus our image encryption scheme can protect the privacy of image content.

After receiving *NER* from cloud, the user will run the *ImageDec(SK, EW)* algorithm to decrypt images, shown as Algorithm 2. The *M* and *N* in Algorithms 1 and 2 are the height and width of images. The KMC also uses this algorithm to decrypt images.

Algorithm 1 ImageEnc(SK, W)

```
while j ≠ M do
    while k ≠ N do
        EW_jk ⇐ SK_{j×N+k} ⊕ W_jk;
        k ⇐ k + 1;
    end while
    j ⇐ j + 1;
end while
```

Algorithm 2 ImageDec(SK, EW)

```
while j ≠ M do
    while k ≠ N do
        W_jk ⇐ SK_{j×N+k} ⊕ EW_jk;
        k ⇐ k + 1;
    end while
    j ⇐ j + 1;
end while
```

5.3.3. ImageFeatureEnc — image feature encryption

In order to preserve the privacy of image features, image features should be encrypted before they are outsourced to the cloud. Image owners and authorized query users both need to encrypt image features; they use the same method to encrypt image features. For an image feature $f_i = \{a_1, a_2, \ldots, a_l\}$, they will first calculate the square of image feature f_i to obtain $f_i{}^2 = \{a_1{}^2, a_2{}^2, \ldots, a_l{}^2\}$ where l is the dimension of image feature f_i. We use the secure multi-party computation method introduced in Section 3.2 to encrypt image features. The *ImageFeatureEnc* process will be described in detail as follows.

First, similar to secure multi-party computation, they choose q as a large primer number, whose length is the same as p, satisfies that q divided by p-1. Then, they select a random number $h \in Z_p$ and generate the g_1 and g_2 as Eqs. (1) and (2).

Second, they randomly choose a number $r_j \in Z_q$ and calculate a number R_j as Eq. (3) for each dimension a_j in image feature f_i.

Finally, they can get the cipertext ea_j of each dimension a_j in f_i by calculating $ea_j = (1+a_jp)R_j \bmod p^2$. Then, the cipertext of $f_i{}^2$ can be obtained by the same way. The encrypted feature and encrypted square feature are shown as follows.

$ef_i = \{ea_1, ea_2, \ldots, ea_l\} = \{(1+a_1p)R_1 \bmod p^2, (1+a_2p)R_2 \bmod p^2, \ldots, (1+a_lp)R_l \bmod p^2\}$.

$ef_i{}^2 = \{ea_1{}^2, ea_2{}^2, \ldots, ea_l{}^2\} = \{(1+a_1{}^2p)R_1 \bmod p^2, (1+a_2{}^2p)R_2 \bmod p^2, \ldots, (1+a_l{}^2p)R_l \bmod p^2\}$.

After encrypting all image features, they can obtain the encrypted feature collection *EF* and *EFF*.

It should be brought to attention that the parameter p is public to all image owners, authorized query users, and cloud. Parameters q, h, and r_i in the image feature encryption process can be selected differently among image features. Image owners and authorized users can select by themselves without communicating with others. Additionally, they do not need to store it, which means these parameters can be discarded after using.

5.4. New approach to manage image similarity

In the field of image retrieval, the similarity of images is always measured by calculating the distance of image features. If two images are similar, then the distance of them will be very small. Euclidean distance is a type of distance that is typically used to measure the similarity of images, shown as Eq. (7). It can calculate the similarity of images accurately. While this will also lead to the problem that when using it to measure the similarity of images in the cloud, the cloud can obtain the image similarity relation information. In order to solve this problem, we propose a new approach to manage image similarity.

Given two image features $X = \{x_1, x_2, \ldots, x_l\}$ and $Y = \{y_1, y_2, \ldots, y_l\}$, the distance *NewDis* between X and Y can be calculated as Eq. (8). Compared with Euclidean distance, we use $\sum_{i=1}^{u} x_i/u$ and $\sum_{i=1}^{u} y_i/u$ to replace x_i and y_i respectively in the third part of Eq. (7).

$$EucDis = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2 + \cdots + (x_u-y_u)^2} = \sqrt{\sum_{i=1}^{u} x_i^2 + \sum_{i=1}^{u} y_i^2 - \sum_{i=1}^{u} 2x_iy_i} \tag{7}$$

$$\begin{aligned} NewDis &= \sqrt{\sum_{i=1}^{u} x_i^2 + \sum_{i=1}^{u} y_i^2 - \sum_{i=1}^{u} 2\frac{\sum_{i=1}^{u} x_i}{u}\frac{\sum_{i=1}^{u} y_i}{u}} \\ &= \sqrt{\sum_{i=1}^{u} x_i^2 + \sum_{i=1}^{u} y_i^2 - 2u\frac{\sum_{i=1}^{u} x_i}{u}\frac{\sum_{i=1}^{u} y_i}{u}} \\ &= \sqrt{\sum_{i=1}^{u} x_i^2 + \sum_{i=1}^{u} y_i^2 - 2\frac{\sum_{i=1}^{u} x_i \sum_{i=1}^{u} y_i}{u}} \end{aligned} \tag{8}$$

Please cite this article in press as: M. Shen, et al., Content-based multi-source encrypted image retrieval in clouds with privacy preservation, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.04.089.

The experimental results show that the accuracy and recall rate of the proposed approach are comparable to the European distance. In addition, the proposed approach can support multi-source encrypted image retrieval. Therefore we used the proposed approach to calculate the similarity between the images in our scheme.

5.5. Secure content-based image retrieval

Before requesting a query, authorized query user should generate a secret query $Q = \{EQ, EQQ, UID, AK\}$, and then send Q to the cloud for image retrieving. After receiving Q from authorized query user, the cloud first verifies whether this user is authorized and which owner authorized this user. If validated successfully, the cloud will retrieve in authorized images in the image database through the retrieval index.

For image collection W and query image collection QW, the similarity of image w_i and image qw_j can be measured by calculating the distance between f_i and qf_j.

However, image collection and image feature collection in the cloud are all encrypted. The similarities between image w_i and image qw_j can be measured by calculating the distance between ef_i and eq_j.

As image features are all encrypted by the secure multi-party computation method, it is the same for the encrypted image feature $ef_i = \{ea_1, ea_2, \ldots, ea_l\}$ and encrypted query image feature $eq_j = \{eqa_1, eqa_2, \ldots, eqa_l\}$ where l is the dimension of image feature. The distance between them can be calculated according to Eqs. (5), (6) and (9). We will describe them as follows.

First, the cloud stores the encrypted image feature $ef_i = \{ea_1, ea_2, \ldots, ea_l\}$ and $ef_i{}^2 = \{ea_1{}^2, ea_2{}^2, \ldots, ea_l{}^2\}$. He will receive the encrypted query image feature $eq_j = \{eqa_1, eqa_2, \ldots, eqa_l\}$ and $eq_j{}^2 = \{eqa_1{}^2, eqa_2{}^2, \ldots, eqa_l{}^2\}$ from an authorized query user. Then, he can get CEA, CEA^2, $CEQA$, $CEQA^2$ according to Eq. (5).

Second, we can obtain the $\sum_{i=1}^{l} ea_i$, $\sum_{i=1}^{l} ea_i^2$, $\sum_{i=1}^{l} eqa_i$ and $\sum_{i=1}^{l} eqa_i^2$, shown as follows.

$$(CEA - 1)/p = \sum_{i=1}^{l} ea_i \mod p$$

$$(CEA^2 - 1)/p = \sum_{i=1}^{l} ea_i^2 \mod p$$

$$(CEQA - 1)/p = \sum_{i=1}^{l} eqa_i \mod p$$

$$(CEQA^2 - 1)/p = \sum_{i=1}^{l} eqa_i^2 \mod p$$

Finally, the distance Sim between ef_i and eq_j can be obtained by Eq. (9). The similarity of images w_i and image qw_j can be measured by this distance value, and the small distance value indicates that they are similar.

$$Sim = \sqrt{\sum_{i=1}^{l} ea_i^2 + \sum_{i=1}^{l} eqa_i^2 - 2\frac{\sum_{i=1}^{l} ea_i \sum_{i=1}^{l} eqa_i}{l}} \quad (9)$$

However, there exists a problem during the distance calculation process. Ciphertexts in encrypted image features are very large and the computation complexity of calculating the sum is high, which will consume too much time. Therefore, constructing a retrieval index is very necessary for high retrieval efficiency.

Table 2
Retrieval index.

Image owner	Image ID	$\sum_{i=1}^{l} ea_i$	$\sum_{i=1}^{l} ea_i^2$
OID_1	$OID_1\_Image_1$	$OID_1\_Image_1\_Sum_1$	$OID_1\_Image_1\_Sum_2$
OID_1	$OID_1\_Image_2$	$OID_1\_Image_2\_Sum_1$	$OID_1\_Image_2\_Sum_2$
OID_2	$OID_2\_Image_1$	$OID_2\_Image_1\_Sum_1$	$OID_2\_Image_1\_Sum_2$
...	...	...	...
OID_n	$OID_n\_Image_1$	$OID_n\_Image_1\_Sum_1$	$OID_n\_Image_1\_Sum_2$

5.6. Index construction

We will calculate the sum of encrypted elements $\sum_{i=1}^{l} ea_i$ and $\sum_{i=1}^{l} ea_i^2$ in encrypted features by using the secure multi-party computation during the distance computation process of image features. However, cipertext ea_i and ea_i^2 in the encrypted feature are very large, which will results in the sum calculation operation having high computation complexity and consuming too much time. Therefore, we should build a retrieval index to improve retrieval efficiency. Because the time mainly consumes in computing the sum of cipertext in encrypted features, given an encrypted image feature $ef_i = \{ea_1, ea_2, \ldots, ea_l\}$, we can calculate $\sum_{i=1}^{l} ea_i$, $\sum_{i=1}^{l} ea_i^2$ in advance and then store them in the retrieval index table, shown as Table 2.

Given an encrypted query image feature $eq_j = \{eqa_1, eqa_2, \ldots, eqa_l\}$, the cloud only needs to calculate $\sum_{i=1}^{l} eqa_i$, $\sum_{i=1}^{l} eqa_i^2$ and then get $\sum_{i=1}^{l} ea_i$, $\sum_{i=1}^{l} ea_i^2$ from index table when executing encrypted image retrieval. The sum of ciphertexts in encrypted feature is computed in advance which will save much time in the retrieval process and the retrieval efficiency of our scheme is improved.

5.7. Images and index update in cloud

Sometimes image owners may add images to the cloud or delete images from the cloud. When, images in the cloud are changed, image features will also be changed. Images in the retrieval index should be in accord with images in the cloud, therefore the retrieval index should be modified when images in the cloud are changed. Our scheme supports the dynamic updating of images and index in the cloud including update, delete, and add operation.

(1) Add operation.
When an image owner requests the cloud to add some images for him, he should send encrypted images and encrypted image features to the cloud. Then, cloud will store these new images and image features in the image database. After that, the cloud will calculate the related data $\sum_{i=1}^{l} ea_i$ and $\sum_{i=1}^{l} ea_i^2$ of encrypted image features and add these two data along with OID, image id into the index table.

(2) Delete operation.
When an image owner wants to delete images in the cloud, he should send the image IDs to the cloud. If there are image IDs that belong to this image owner, the cloud will delete these images from the index table. The cloud will also delete encrypted images and the corresponding encrypted image features.

(3) Update operation.
For preserving the image privacy, an image owner may re-encrypt his images, image features, and then update these images in the cloud. When an image owner requests to update images, the cloud should delete the stale encrypted images and image features using the image IDs. Then, the cloud adds these re-encrypted images and their image features into the image database. Since updating encrypted image features will not change the $\sum_{i=1}^{l} ea_i$ and $\sum_{i=1}^{l} ea_i^2$, there is no need to update the index table.

Please cite this article in press as: M. Shen, et al., Content-based multi-source encrypted image retrieval in clouds with privacy preservation, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.04.089.

6. Security analysis

In this section, we will analyze the security of our scheme, including data privacy and image similarity leakage in the cloud.

6.1. Data privacy

In our scheme, data privacy contains the privacy of image content, image features and image similarity information in cloud. We will analyze these three kinds of privacy in the following subsections.

6.1.1. Image content privacy

As described in Section 5.2, the image owner generates a secret key *SK* to encrypt images. The image owner does not want unauthorized user (e.g., cloud, adversary or others) to obtain his image content; he will not reveal his secret key to unauthorized user. In our scheme, the image owner needs to store his secret key in the KMC. We assume the KMC is fully trusted and will not reveal secret keys to unauthorized user. At the same time, we assume the authorized query user will not reveal his secret image encryption key to unauthorized user and will not send image retrieval results to unauthorized user. For the privacy of images outsourced in the cloud, our scheme supports dynamically updating images in the cloud, which means that image owners can re-encrypt images and then outsource these re-encrypted images to the cloud to replace the stale encrypted images. This operation further enhances the privacy protection of images in the cloud. Once an unauthorized user obtains the key of an image owner, he can only crack images of this image owner for a certain period of time, and if the image owner updates his encrypted images in the cloud, this unauthorized user will be unable to crack these re-encrypted images. For the privacy of the image retrieval results, the authorized query user should send a secret image encryption key that is different from his previous query to the KMC for each of his query. Once an unauthorized user obtain the secret image encryption of an authorized query user, he can only crack retrieval results of this user for this query time which enhance the image privacy protection. Since image owners, authorized query users and the KMC will not reveal the secret image encryption key to others, unauthorized users are unable to obtain secret image encryption keys and they are also unable to obtain plain image content through image owners, authorized query users or the KMC.

In the existing researches, the cloud returns encrypted retrieval results to authorized query users directly. If we also adopt this strategy, once the unauthorized user obtains one image owner's secret image encryption key, then he can brute force encrypted image retrieval results to obtain the plain image contents. This operation is valid through every query operation of different authorized query users. Even though we have illustrated that the unauthorized user is unable to obtain secret image encryption keys, this strategy may also contain some insecure aspects. Therefore, we designed a different strategy. The cloud first sends *ER* to the KMC after it finishing the retrieval operation. The KMC will decrypt all images received from the cloud and encrypt these images with the secret image encryption key of the related authorized query user, and then send these re-encrypted images *NER* to the cloud. Finally, the cloud sends *NER* that are returned by the KMC to the authorized query user. This new strategy can solve the above problem. Images returned by the cloud are encrypted by the secret image encryption key of the authorized query user, even an unauthorized user obtains secret image encryption keys of image owners, he is unable to crack these images correctly. Even though the unauthorized user obtains the secret image encryption key of an authorized query user, he can only access to crack images that are returned to this authorized query user in this query. Furthermore, the authorized query user will send a different image encryption key to the KMC for each query. Even the unauthorized user can crack retrieval results of this query, he is unable to crack retrieval results in the next query using the same key. In our scheme, the unauthorized users are unabled to obtain secret keys of image owners and authorized query users, therefore our scheme can guarantee that image content privacy is unable to be captured by unauthorized user.

6.1.2. Image feature privacy

We use the sum protocol in secure multi-party computation model proposed in [25] to encrypt image features. Jung et al. [25] proposed three security models in his paper, shown as Definitions 1–3.

Definition 1 (*CDH problem in G*)**.** The Computational Diffie–Hellman problem in a multiplicative group G with generator g is defined as follows: given only $g, g^a, g^b \in G$ where $a, b \in Z$, compute g^{ab} without knowing a or b.

Definition 2 (*DDH problem in G*)**.** The Computational Diffie–Hellman problem in a multiplicative group G with generator g is defined as follows: given only $g, g^a, g^b, g^c \in G$ where $a, b, c \in Z$, decide if $g^{ab} = g^c$.

Definition 3 (*CDH-Security in G*)**.** We say our privacy preserving (sum or product) calculation is CDH-secure in G if any Probabilistic Polynomial Time Adversary (PPTA) who cannot solve the CDH problem with non-negligible chance has negligible chance to infer any honest participants private value in G, i.e., any PPTAs probability to solve the CDH problem ϵ satisfies $\epsilon < |1/p(\kappa)|$ for any polynomial p$(\cdot)$ where κ is the order of the group G defined in the CDH problem.

In Jung's paper, each participant P_i will receive the $g_2^{r_{i-1}}$ and $g_2^{r_{i+1}}$ sent from participant P_{i-1} and P_{i+1}; therefore, the unauthorized user may obtain $g_2^{r_{i-1}}$, $g_2^{r_{i+1}}$ and $g_2^{r_i}$. They has proved that their sum protocol is CDH-secure in G_2 in this condition.

Theorem 6.1. *Our scheme can protect image feature privacy from being captured by cloud and unauthorized users.*

Proof. In our scheme, image owner generates $g_2^{r_i}$ for each dimension. The cloud and unauthorized users are enabled to obtain $g_2^{r_{i-1}}$, $g_2^{r_{i+1}}$ and $g_2^{r_i}$, thus our image feature privacy is also CDH-secure in G_2.

For an image feature $f_i = \{a_1, a_2, \ldots, a_l\}$, the image owner will calculate the $f_i{}^2 = \{a_1{}^2, a_2{}^2, \ldots, a_l{}^2\}$ and encrypt them to obtain ef_i and $ef_i{}^2$. The cipertext a_i and $a_i{}^2$ are shown as follows:

$$ea_i = (1 + a_i p)R_i \mod p^2 = (1 + a_i p)g_2^{(r_{i+1}/r_{i-1})r_i} \mod p^2$$

$$ea_i^2 = (1 + a_i^2 p)R_i \mod p^2 = (1 + a_i^2 p)g_2^{(r_{i+1}/r_{i-1})r_i} \mod p^2$$

If the unauthorized user wants to obtain a_i, then he has to solve the secret parameter R_i. Because the calculation process of R_i is unknown to him, he is unable to solve R_i, and he cannot obtain the plaintext a_i.

At the same time, the parameter collection $PC = \{g_2, \{r_i\}\}$ can be chosen among different image features. This means if we want to encrypt image feature f_1, f_1^2, f_2, f_2^2, we can choose four different

Please cite this article in press as: M. Shen, et al., Content-based multi-source encrypted image retrieval in clouds with privacy preservation, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.04.089.

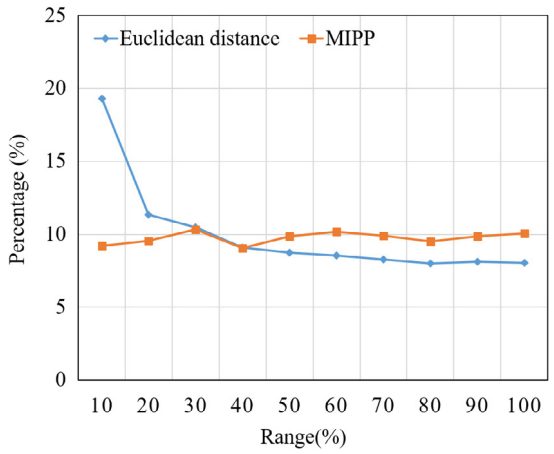


Fig. 2. The percentage that images will appear in each range of all returned images.

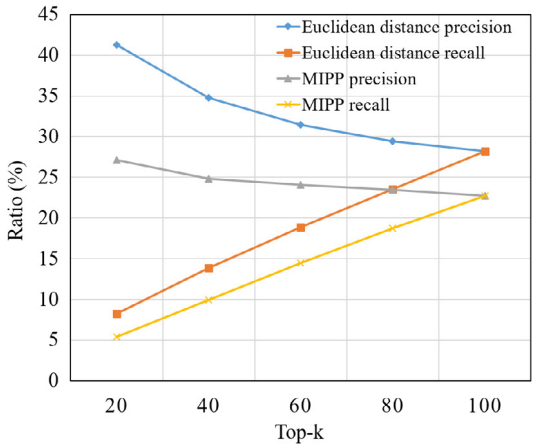


Fig. 3. Retrieval accuracy of our scheme and Euclidean distance scheme.

parameter collections *PC* to encrypt them. Even if unauthorized users obtain one parameter collection *PC*, he is only able to decrypt one image feature. However, unauthorized users are unable to get the parameter collection *PC*. Moreover, we can discard the parameter collection *PC* directly, as there is no need to store it.

Furthermore, our scheme supports the update of image features in cloud, which enables image owners update their encrypted image features at any time. Once an unauthorized user obtains one parameter collection *PC*, he can only crack one or more encrypted image features that are encrypted by this parameter collection *PC* and if image owners update their encrypted image features in the cloud then the parameter collection *PC* that this unauthorized user obtains will be invalidation. This unauthorized user need to obtain the new parameter collection *PC* to crack encrypted image features. However, image owners will not store the parameter collection *PC* and they will not reveal the *PC* to unauthorized users, so unauthorized users will not obtain the *PC* and they will be unable to crack the plaintext of image features.

According to the above description, unauthorized users are unable to obtain the plaintext of image features. Thus, our scheme can protect the image feature privacy from being captured by unauthorized users.

6.2. Image similarity leakage in cloud

In our scheme, we assume that the cloud is "honest but curious", which means that he will execute the image retrieval operation accurately and at the same time he will analyst the relation or other information of images. In current research, there always exists the image similarity leakage problem in the cloud during the cloud executing image retrieval operation. Images in the retrieval result are arranged by the similarity to query image, which will reveal the similarity information of images to the cloud.

A new distance to manage the similarity of images is proposed in our scheme, which can solve the above problem. Fig. 2 shows the distribution of images that is similar to query images in retrieval results when the cloud returns top 100 images. The abscissa is the percentage that similar images distribute in the retrieval results, and the ordinate is the probability that similar images distribute in the designated range. We can see that when retrieving with the Euclidean distance, the percentage of truly similar images appear in the top 10% of all retrieval result images, which is very high, and the distribution percentage is decreasing from beginning to end in retrieval results. The truly similar image distribution of our scheme is uniform and similar images are not likely to distribute at the beginning percentage of the retrieval result, which will mislead the cloud to analyst the similarity of the images. Because similar images are uniformly distributed in the retrieval results, the cloud may get the wrong similarity relation. Therefore, our scheme can prevent the image similarity leakage to the cloud.

7. Performance evaluation

In this section, we will introduce the performance evaluation of our scheme, including experimental setting, evaluation of retrieval accuracy, evaluation of time consumptions, and evaluation of storage consumption.

7.1. Experimental setting

The corel images [26,27] data set is usually used to verify the experiment scheme in the research field of image retrieval. It contains 100 categories each of which has 100 images. They are selected as test images in our experiments. We choose 10 categories and generate 5 queries for each category, so there are 50 queries in total to evaluate the retrieval accuracy. The proposed scheme is implemented by C++on Intel Core(TM) Processor 2.7 GHz.

7.2. Evaluation of retrieval accuracy

In the field of information retrieval, precision, recall ratio, and F1-Measure are typical metrics to evaluate the retrieval results as formulated in Eqs. (10)–(12), where TP represents the true positives, FP represents the false positives, and FN represents the false negatives. The precision and recall always represent the contrary variation tendency as shown in Fig. 3. We can use the F1-Measure to make a comprehensive evaluation. Fig. 3 shows that the retrieval accuracy of our scheme is about 10% lower than the scheme retrieval by Euclidean distance on average. The retrieval recall of our scheme is about 5% lower than the scheme retrieval by Euclidean distance on average. Fig. 4 shows the F1-Measure of MIPP scheme and the scheme retrieving with Euclidean distance. The result shows the F1-Measure of our scheme is about 7% lower than the scheme retrieving with Euclidean distance. As described in previous sections, our scheme supports multiple image owners and can preserve image similarity information in the cloud. Therefore, the loss of retrieval accuracy of our scheme is the trade off these two aspects.

$$P = \frac{TP}{TP + FP} \tag{10}$$

Please cite this article in press as: M. Shen, et al., Content-based multi-source encrypted image retrieval in clouds with privacy preservation, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.04.089.

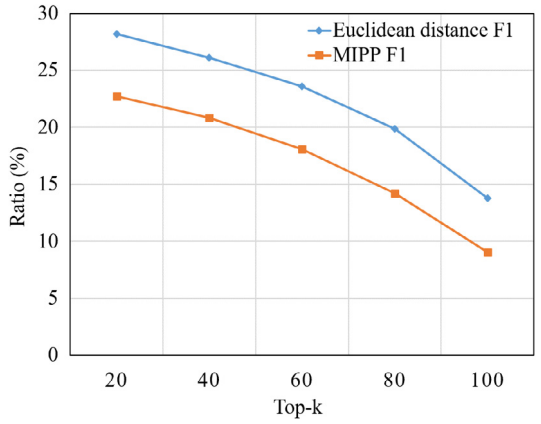


Fig. 4. F1-Measure of our scheme and Euclidean distance scheme.

$$R = \frac{TP}{TP + FN} \tag{11}$$

$$F1 = \frac{2PR}{P + R} \tag{12}$$

7.3. Evaluation of time consumptions

The time consumptions of our scheme primarily consist of index construction time and secure image retrieval time, which is described as follows:

(1) Index construction time.
When a new image owner participates in our scheme, the image size will be increased. Fig. 5 shows the index construction time is increasing with the larger size of images. When the image size is 10,000, the index construction time is approximately 7 min, which is tolerable. After constructing the retrieval index, the efficiency of secure image retrieval in our scheme can be improved.

(2) Secure image retrieval time.
The time consumptions of plain image retrieval, encrypted image retrieval with index, and encrypted image retrieval without index are shown in Figs. 6–8 respectively. The results show that the larger image size, the more time image retrieval consumes. The plain image retrieval time of 10,000 images consumes approximately 1 s. Encrypted image retrieval time without index of 10,000 images consumes approximately 6.8 min, while encrypted image retrieval time with index of 10,000 images consumes approximately 50 ms. According to the above experimental data, we can calculate that index-based image retrieval takes approximately 8,160 times faster than non-indexed image retrieval and index-based image retrieval takes approximately 1,200 times faster than plain image retrieval when retrieving in a collection of 10,000 images. We can conclude that index-based encrypted image retrieval can greatly improve retrieval efficiency compared with plain image retrieval and non-indexed encrypted image retrieval. The retrieval time-consuming result of index-based encrypted image retrieval shows that the retrieval efficiency of our scheme is appreciable.

Table 3
Storage consumption of 10000 images.

	Encrypted image features	Retrieval index
Storage consumption (KB)	3352278	267

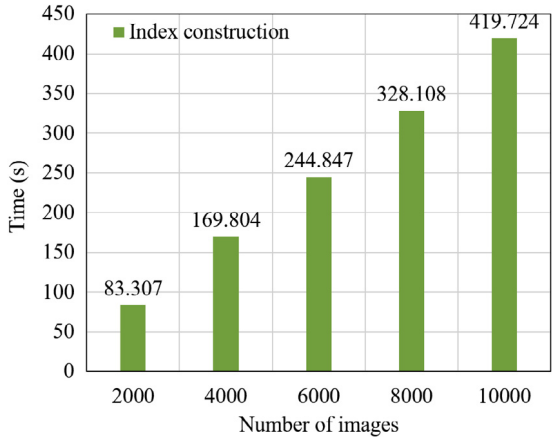


Fig. 5. Index construction consumption.

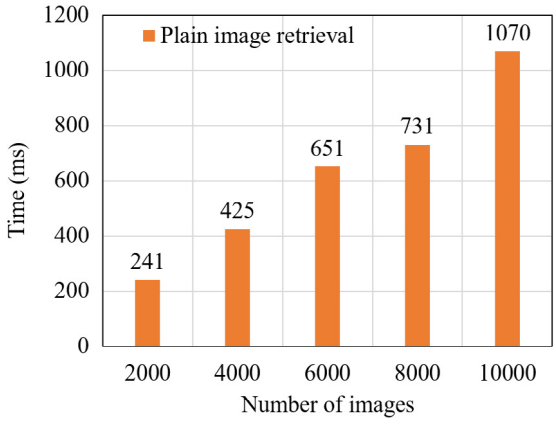


Fig. 6. Plain image retrieval consumption.

7.4. Evaluation of storage consumption

The storage consumption includes index storage and encrypted image features storage consumption, described as follows:

(1) Index storage consumption.
We extract image features to represent images for similarity calculation and construct retrieval indexes for improving retrieval efficiency. The storage consumption of index table is shown in Table 3. We can find that the retrieval index of 10,000 images costs approximately 267 KB storage space. Therefore, the storage consumption of index table is very low.

(2) Encrypted image features storage consumption.
In order to preserve image features privacy, we outsource encrypted image features to cloud. Table 3 shows that 10,000 encrypted image features consume approximately 3.2 GB storage space. Because encrypted image features are stored in the cloud, which has high storage facilities, the storage consumption of encrypted image features in our scheme is tolerable.

Please cite this article in press as: M. Shen, et al., Content-based multi-source encrypted image retrieval in clouds with privacy preservation, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.04.089.

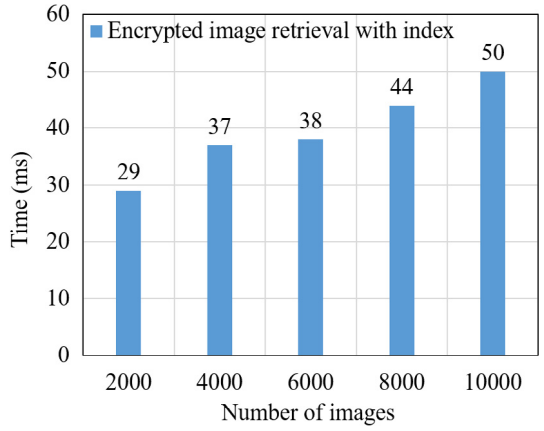


Fig. 7. Consumption of encrypted image retrieval with index.

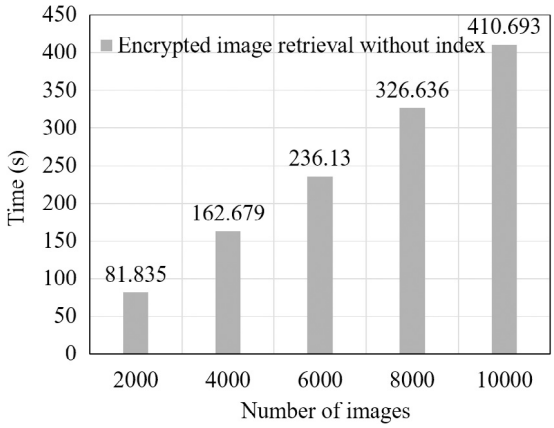


Fig. 8. Consumption of encrypted image retrieval without index.

8. Conclusion

In this paper, we presented a content-based multi-source encrypted image retrieval scheme in clouds with privacy protection. We encrypted image features with the secure multi-party computation, which allowed image owners to encrypt image features by using their own keys. We also proposed a new method to measure the similarity of images that could avoid revealing image similarity information to the cloud at a certain extent. Theoretical analysis and experimental results showed that our scheme enabled an accurate and efficient image retrieval over images gathered from multiple sources, while providing privacy guarantees. In the future work, we are to further improve the image retrieval efficiency.

Acknowledgments

This work was supported in part by the National Science Foundation of China [Grant number 61602039] and the China National Key Research and Development Program [Grant number 2016YFB0800301].

References

[1] M. Kaur, N. Sohi, A novel technique for content based image retrieval using color, texture and edge features, in: 2016 International Conference on Communication and Electronics Systems, ICCES, 2016, pp. 1–7. http://dx.doi.org/10.1109/CESYS.2016.7889955.

[2] A. Yalavarthi, K. Veeraswamy, K.A. Sheela, Content based image retrieval using enhanced gabor wavelet transform, in: 2017 International Conference on Computer, Communications and Electronics, Comptelix, 2017, pp. 339–343. http://dx.doi.org/10.1109/COMPTELIX.2017.8003990.

[3] A. Rashno, S. Sadri, Content-based image retrieval with color and texture features in neutrosophic domain, in: 2017 3rd International Conference on Pattern Recognition and Image Analysis, IPRIA, 2017, pp. 50–55. http://dx.doi.org/10.1109/PRIA.2017.7983063.

[4] Y. Chen, The image retrieval algorithm based on color feature in: 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), 2016, pp. 647–650. http://dx.doi.org/10.1109/ICSESS.2016.7883151.

[5] C.H. Su, H.S. Chiu, T.M. Hsieh, An efficient image retrieval based on HSV color space in: 2011 International Conference on Electrical and Control Engineering, 2011, pp. 5746–5749. http://dx.doi.org/10.1109/ICECENG.2011.6058026.

[6] S.S. Devi, R. Balasundaram, Content based texture image retrieval based on modified dominant directional local binary pattern, in: 2017 4th International Conference on Advanced Computing and Communication Systems, ICACCS, 2017, pp. 1–6. http://dx.doi.org/10.1109/ICACCS.2017.8014592.

[7] X. Chen, Y. Zheng, C. Yu, C. Gao, Image retrieval based on color and texture features in: 2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2013, pp. 403–406. http://dx.doi.org/10.1109/IIH-MSP.2013.107.

[8] A.K. Naveena, N.K. Narayanan, Image retrieval using combination of color, texture and shape descriptor, in: 2016 International Conference on Next Generation Intelligent Systems, ICNGIS, 2016, pp. 1–5. http://dx.doi.org/10.1109/ICNGIS.2016.7854023.

[9] A. Anandh, K. Mala, S. Suganya, Content based image retrieval system based on semantic information using color, texture and shape features, in: 2016 International Conference on Computing Technologies and Intelligent Data Engineering, ICCTIDE'16, 2016, pp. 1–8. http://dx.doi.org/10.1109/ICCTIDE.2016.7725364.

[10] M. Shen, B. Ma, L. Zhu, R. Mijumbi, X. Du, J. Hu, Cloud-based approximate constrained shortest distance queries over encrypted graphs with privacy protection, IEEE Trans. Inf. Forensics Secur. 13 (4) (2018) 940–953. http://dx.doi.org/10.1109/TIFS.2017.2774451.

[11] L. Zhu, X. Tang, M. Shen, X. Du, M. Guizani, Privacy-preserving DDoS attack detection using cross-domain traffic in software defined networks, IEEE J. Sel. Areas Commun. (2018). http://dx.doi.org/10.1109/JSAC.2018.2815442. 1–1.

[12] Y. Zhang, L. Zhuo, Y. Peng, J. Zhang, A secure image retrieval method based on homomorphic encryption for cloud computing, in: 2014 19th International Conference on Digital Signal Processing, 2014, pp. 269–274. http://dx.doi.org/10.1109/ICDSP.2014.6900669.

[13] L. Zhang, T. Jung, K. Liu, X.Y. Li, X. Ding, J. Gu, Y. Liu, Pic: Enable large-scale privacy preserving content-based image search on cloud, IEEE Trans. Parallel Distrib. Syst. 28 (11) (2017) 3258–3271. http://dx.doi.org/10.1109/TPDS.2017.2712148.

[14] Z. Xia, Y. Zhu, X. Sun, Z. Qin, K. Ren, Towards privacy-preserving content-based image retrieval in cloud computing, IEEE Trans. Cloud Comput. 6 (1) (2018) 276–286. http://dx.doi.org/10.1109/TCC.2015.2491933.

[15] J. Yuan, S. Yu, L. Guo, Seisa: Secure and efficient encrypted image search with access control, in: 2015 IEEE Conference on Computer Communications, INFOCOM, 2015, pp. 2083–2091. http://dx.doi.org/10.1109/INFOCOM.2015.7218593.

[16] X. Yuan, X. Wang, C. Wang, A.C. Squicciarini, K. Ren, Towards privacy-preserving and practical image-centric social discovery, IEEE Trans. Dependable Secure Comput. (2017). http://dx.doi.org/10.1109/TDSC.2016.2609930. 1–1.

[17] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun, K. Ren, A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing, IEEE Trans. Inf. Forensics Secur. 11 (11) (2016) 2594–2608. http://dx.doi.org/10.1109/TIFS.2016.2590944.

[18] L. Weng, L. Amsaleg, T. Furon, Privacy-Preserving outsourced media search, IEEE Trans. Knowl. Data Eng. 28 (10) (2016) 2738–2751. http://dx.doi.org/10.1109/TKDE.2016.2587258.

[19] B. Ferreira, J. Rodrigues, J. Leitão, H. Domingos, Privacy-preserving content-based image retrieval in the cloud, in: 2015 IEEE 34th Symposium on Reliable Distributed Systems, SRDS, 2015, pp. 11–20. http://dx.doi.org/10.1109/SRDS.2015.27.

[20] X. Zhang, H. Cheng, Histogram-based retrieval for encrypted jpeg images in: 2014 IEEE China Summit International Conference on Signal and Information Processing, ChinaSIP, 2014, pp. 446–449. http://dx.doi.org/10.1109/ChinaSIP.2014.6889282.

[21] H. Cheng, X. Zhang, J. Yu, F. Li, Markov process based retrieval for encrypted jpeg images, in: 2015 10th International Conference on Availability, Reliability and Security, 2015, pp. 417–421. http://dx.doi.org/10.1109/ARES.2015.18.

[22] C. Zhang, J. Li, S. Wang, Z. Wang, An encrypted medical image retrieval algorithm based on DWT-DCT frequency domain, in: 2017 IEEE 15th International

Conference on Software Engineering Research, Management and Applications, SERA, 2017, pp. 135–141. http://dx.doi.org/10.1109/SERA.2017.7965719.

[23] B.S. Manjunath, J.R. Ohm, V.V. Vasudevan, A. Yamada, Color and texture descriptors, IEEE Trans. Circuits Syst. Video Technol. 11 (6) (2001) 703–715. http://dx.doi.org/10.1109/76.927424.

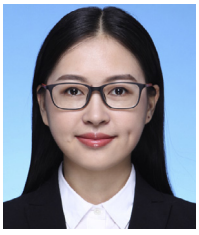
[24] M. Mejía-Lavalle, C.P. Lara, J.R. Ascencio, The MPEG-7 visual descriptors: A basic survey in: 2013 International Conference on Mechatronics, Electronics and Automotive Engineering, 2013, pp. 115–120. http://dx.doi.org/10.1109/ICMEAE.2013.46.

[25] T. Jung, X.Y. Li, M. Wan, Collusion-tolerable privacy-preserving sum and product calculation without secure channel, IEEE Trans. Dependable Secure Comput. 12 (1) (2015) 45–57. http://dx.doi.org/10.1109/TDSC.2014.2309134.

[26] J. Li, J.Z. Wang, Automatic linguistic indexing of pictures by a statistical modeling approach, IEEE Trans. Pattern Anal. Mach. Intell. 25 (9) (2003) 1075–1088. http://dx.doi.org/10.1109/TPAMI.2003.1227984.

[27] J.Z. Wang, J. Li, G. Wiederhold, Simplicity: Semantics-sensitive integrated matching for picture libraries, IEEE Trans. Pattern Anal. Mach. Intell. 23 (9) (2001) 947–963. http://dx.doi.org/10.1109/34.955109.

Meng Shen received the B.Eng degree from Shandong University, Jinan, China in 2009, and the Ph.D. degree from Tsinghua University, Beijing, China in 2014, both in computer science. Currently he serves in Beijing Institute of Technology, Beijing, China, as an assistant professor. His research interests include privacy protection of cloud-based services, network virtualization and traffic engineering. He received the Best Paper Runner-Up Award at IEEE IPCCC 2014. He is a member of the IEEE.



Guohua Cheng is a graduate student in the School of Computer Science, Beijing Institute of Technology. Her research interests include image fusion, image retrieval, and privacy preserving algorithms.

Liehuang Zhu is a professor in the Department of Computer Science at Beijing Institute of Technology. He is selected into the Program for New Century Excellent Talents in University from Ministry of Education, P.R. China. His research interests include Internet of Things, Cloud Computing Security, Internet and Mobile Security.

Xiaojiang Du is a tenured professor in the Department of Computer and Information Sciences at Temple University, Philadelphia, USA. His research interests are wireless communications, wireless networks, security, and systems. He has authored over 230 journal and conference papers in these areas, as well as a book published by Springer. Dr. Du has been awarded more than $5 million US dollars research grants from the US National Science Foundation (NSF), Army Research Office, Air Force Research Lab, NASA, the State of Pennsylvania, and Amazon. He serves on the editorial boards of three international journals. Dr. Du is a Senior Member of IEEE and a Life Member of ACM.

Jiankun Hu is a Professor at the School of Engineering and IT, University of New South Wales (UNSW) Canberra (also named UNSW at the Australian Defence Force Academy (UNSW@ADFA), Canberra, Australia). He is the invited expert of Australia Attorney-Generals Office assisting the draft of Australia National Identity Management Policy. Prof. Hu has served at the Panel of Mathematics, Information and Computing Sciences (MIC), ARC ERA (The Excellence in Research for Australia) Evaluation Committee 2014. His research interest is in the field of cyber security covering intrusion detection, sensor key management, and biometrics authentication. He has many publications in top venues including IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Computers, IEEE Transactions on Parallel and Distributed Systems (TPDS), IEEE Transactions on Information Forensics & Security (TIFS), Pattern Recognition, and IEEE Transactions on Industrial Informatics. He is the associate editor of the IEEE Transactions on Information Forensics and Security.

Please cite this article in press as: M. Shen, et al., Content-based multi-source encrypted image retrieval in clouds with privacy preservation, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.04.089.