NavaJyoti, International Journal of Multi-Disciplinary
Research Volume 3, Issue 1, August 2018

Privacy-Preserving Content-Based Image Retrieval in Cloud

Nancy Esther Devakumar, Vinoliya Grace.E, Irene Getzi

[1]*Department MCA, Jyoti Nivas College, Bangalore, India*
[2] *Department MCA, Jyoti Nivas College, Bangalore, India*
[3] *Department MCA, Jyoti Nivas College, Bangalore, India*

ABSTRACT

Cloud storage systems are increasingly being used to host personal or organizational data of critical importance, especially for image data that needs more storage space than ordinary data. While bringing in much convenience, existing cloud storage solutions could seriously breach the privacy of users. Encryption before outsourcing images to the cloud helps to protect the privacy of the data, but it also brings challenges to perform image retrieval over encrypted data. In this paper, we propose a framework for the privacy-preserving outsourced storage, search, and retrieval of images in large-scale, dynamically updated repositories. Our framework is composed of two main components: an image encryption component, executed on client devices; and a storage, indexing, and searching component (in the encrypted domain), executed in the outsourcing server (e.g. a cloud provider). We base this framework on a new encryption scheme specifically designed for images, called IES-CBIR.

INTRODUCTION

With the development of the imaging devices, such as digital cameras, smartphones and medical imaging equipment, our world has been witnessing a tremendous growth in quantity, availability, and importance of images. The needs of efficient image storage and retrieval services are reinforced by the increase of large-scale image databases among all kinds of areas. Compared with text documents, images consume much more storage space. Hence, its maintenance is considered to be a typical example for cloud storage outsourcing. Content based search and retrieval solutions is the need for the availability of large amounts of images in public and private storage systems. Data outsourcing in Cloud Computing may seem like a natural solution to support large scale storage and image retrieval but there are new challenges concerning data control and privacy. For privacy-preserving purposes, sensitive images, such as medical and personal images, need to be encrypted before outsourcing. In order to secure the data in cloud, the proposed system supports CBIR over encrypted images without leaking the sensitive information to the cloud server.

Recent news has provided clear proofs that privacy should not be expected to be preserved from Cloud providers. Furthermore, malicious system administrators working for the providers have full access to data on the hosting cloud machines. Finally, external hackers can exploit software vulnerabilities to gain unauthorized access to servers. The traditional approach to address privacy is to encrypt sensitive data before outsourcing and run all computations on the client side. But, by doing this, it imposes a lot of overload on client as the data must be continuously be downloaded, decrypted, processed and securely re-uploaded. Many applications such as online and mobile applications that operate over large datasets cannot cope with this overload. A more feasible solution would be to outsource computations and perform operations over the encrypted data on the server side.

To address these challenges, we propose a new secure framework for the privacy-preserving outsourced storage, search, and retrieval of large-scale, dynamically updated image repositories. We base our proposal on another contribution of this work: IES-CBIR, a novel Image Encryption Scheme with Content-Based Image Retrieval properties. Key to the design of our scheme is the observation that in images, color information can be separated from texture information, enabling the use of different encryption techniques with different properties for protecting each of these features. Based on this observation, we consider texture more relevant to the color in object recognition.

Previous approaches for supporting outsourced storage, search and retrieval of images in the encrypted domain can be broadly divided in two classes: Searchable Symmetric Encryption (SSE) based approaches and Public-Key Partially-Homomorphic approaches (PKHE). In SSE-based solutions, clients process and encrypt their data before outsourcing it to Cloud. After this processing, an index is created, encrypted and stored in Cloud which allows clients to search their data efficiently in a secure way. Data is typically encrypted with a probabilistic symmetric-key encryption scheme, while the index is protected through a combination of probabilistic and deterministic encryption. SSE-based approach has the following limitation: (i) Clients either require a trusted proxy or have to index their images and then encrypt that index which involves the use of additional computational power on the client side. (ii) Instead of just uploading images, clients will also have to retrieve and re-upload their encrypted index with each repository update which leads to additional bandwidth usage that has a negative impact on storage operations.

The alternatives to SSE are based on public-key partially-homomorphic encryption (PKHE) schemes. In this approach, clients encrypt their images pixel by pixel with a PKHE scheme, allowing the cloud to process and index their encrypted images on their behalf. By doing this, it avoids many of the practical issues of SSE-based solutions. The limitation of PKHE is it works with much higher time and space complexities.

PROPOSED SCHEME

The main aim of IES-CBIR is to preserve the privacy. Image privacy is defined as the ability to keep the contents of an image secret to unauthorized users. The main objective is to ensure privacy of users, hence all data sent to the cloud is encrypted. Image content refers to the combination of color and texture information. This information helps to identify objects, people etc. in an image. Image privacy can be protected by preventing unauthorized users from recognizing those objects in the image. Further, image color and texture information can be separated from each other. Color information is given from pixel color values in the different channels of some color models while texture information is given by the relative position of pixels and strong color changes across neighboring pixels. Texture information is usually more relevant in images for object recognition. No subcomponent alone i.e. color or texture information can be used to assume the precise contents of an image since the color information is usually uncertain and texture information depends on both the pixel positions and their color values.

Algorithm:

Step 1 - *Generation of Repository Key and Image Key*: IES-CBIR works with two different types of cryptographic keys, repository keys(*rk*) and image keys(*ik*). For the implementation of the repository key, we first generate a random permutation. These permutations can be generated through a Pseudo-Random Generator (*PRG*) *G*, parameterized with some random seed (in our implementation we instantiate G with an AES-based PRG). This results in 3 sub-keys: *rkH*, *rkS*, *rkV*. The range [0...100] represents all possible color values in the HSV color space ((H) hue, (S) saturation, (V) value/brightness), and each different sub-key is attributed to a color channel. Image Key(*ik*) is generated by requesting 128 (*spik*) pseudorandom bits to G. ik will be used as a cryptographic seed for the probabilistic encryption counterpart of IES-CBIR.

Step 2 – ***Encryption:*** Image encryption in IES-CBIR is achieved through two main steps and a final optional step: i) pixel color values encryption, ii) pixel positions permutation and iii) image compression. The goal of the first step is to protect image color features, by applying the Pseudo-Random Permutation (PRP) P on all pixel color values. We choose a specific color-domain PRP, which allows us to preserve the format of encrypted images.

Our construction encrypts pixel color values by deterministically replacing them, in each color channel, using repository key rk = {*rkH*, *rkS*, *rkV*}. The below given equation represents this operation, where *Prk*(*px*) is the encryption of pixel p's value in color component *x* through *P* and key *rkx*.

$$CI \leftarrow Prkx\ (px) : \forall x \in (H, S, V), \forall p \in I$$

This step of encryption securely hides the color values of the encrypted pixels. To fully protect image contents, we rely on a second probabilistic step in our encryption algorithm: (pseudo)random pixel position permutation, through pixel rows and columns shifting. This step consists in the following: a PRG G is instantiated with a previously generated image key ik (operation GENIK above) as cryptographic seed. Then, for each pixel column, we request from G a new pseudorandom value r between 1 and the image height, and do a downward shift on that column of r positions, overflowing to its beginning. After all columns have been randomly shifted, we repeat the procedure for the rows (with random values ranging between 1 and the image width). The below given equations describe this step, where w and h are, respectively, the width and height of image I. Note that this encryption algorithm has no cipher text expansion (i.e., after encryption the image has the same width and height as before).

$$CI\ (x, y) \leftarrow CI\ (x,(y + r) \bmod h) : \forall x \in \{1, .., w\}, \forall y \in \{1, .., h\}$$
$$CI\ (x, y) \leftarrow CI\ ((x + r) \bmod w, y) : \forall x \in \{1, .., w\}, \forall y \in \{1, .., h\}$$

The final, optional step in our encryption algorithm is to perform image compression. This is possible due to the format-preserving properties of IES-CBIR, and can be achieved through the use of a non-lossy image compression scheme such as PNG, directly over the encrypted image.

Step 3 – Decryption: The decryption algorithm applies the different steps of encryption in the opposing order.

$$CI\ ((x + r) \bmod w, y) \leftarrow CI\ (x, y) : \forall x \in \{1, .., w\}, \forall y \in \{1, .., h\}$$
$$CI\ (x,(y + r) \bmod h) \leftarrow CI\ (x, y) : \forall x \in \{1, .., w\}, \forall y \in \{1, .., h\}$$
$$I \leftarrow Prkx(cpx) : \forall x \in (H, S, V), \forall cp \in CI$$

Step 4 - Searching-Trapdoor Generation: The TRPGEN algorithm generates searching trapdoors that users can leverage to search over image repositories. Trapdoor generation requires a query image Q as input, as well as the repository key rk. This means that users with access to rk will be able to access color values of all images stored in that repository. However, users can't access texture information (and hence full image contents) without the corresponding image keys, and can't use rk to search other repositories. Given rk, the TRPGEN algorithm operates in a similar fashion to the ENC algorithm. In the below given equation, image key is substituted by a new randomly generated ik. This means that searching trapdoors are also decryptable, and can be stored in the repositories as new images as long as the querying users locally save the generated image keys.

CBIR in Encrypted Domain

On the cloud's side, the received encrypted images are processed and indexed for CBIR before being stored. Encrypted Image Processing has two main steps: Feature Extraction and Feature Indexing. Feature Extraction consists in processing an image and extracting a reduced set of feature vectors that describe it. We focus on color features in the HSV color model and their representation as color histograms. For each encrypted image and each HSV color channel, the cloud server builds a color histogram by counting the number of pixels in each intensity level. Once these features are extracted, the cloud performs Feature Indexing to speed up the query execution. For this, we use Bag-Of-Visual-Words (BOVW) representation to build a vocabulary tree and an inverted list index for each repository. This approach for indexing shows good search performance and scalability properties.

In the BOVW model, feature-vectors are hierarchically clustered (for instance, using the k-means algorithm) into a vocabulary tree (also known as codebook).

After the creation of the codebook, additional images can be stored dynamically by hierarchically stemming them against it. After processing and indexing encrypted images, the cloud server can receive search requests from users, through the submission of searching trapdoors for some query images of their choice. When a new searching trapdoor is received, the cloud server extracts its color feature-vectors and finds their closest visual words by stemming them against the codebook. The BOVW approach guarantees that only the most relevant images have to be compared in the scoring step ensuring scalability. After receiving these ranked results, users can explicitly request full access to images by requesting corresponding image keys from their owners.

CONCLUSION

The proposed new secure framework IES-CBIR for the privacy-preserving outsourced storage, search, and retrieval of large-scale, dynamically updated image repositories, where the client overload is reduced. Key to the design of IES-CBIR observation that in images, color information can be separated from texture information, enabling the use of different encryption techniques and allowing privacy-preserving Content-Based Image Retrieval to be performed by third-party, untrusted cloud servers.

REFERENCES

[1] Global Web Index, “Instagram tops the list of social network growth,” http://blog.globalwebindex.net/instagram-tops-list-of-growth, 2013.
[2] C. D. Manning, P. Raghavan, and H. Schutze, ¨ An Introduction to Information Retrieval. Cambridge University Press, 2009, vol. 1.
[3] R. Chow, P. Golle, M. Jakobsson, E. Shi, J. Staddon, R. Masuoka, and J. Molina, “Controlling data in the cloud: outsourcing computation without outsourcing control,” in CCSW’09, 2009.
[4] D. Lewis, “iCloud Data Breach: Hacking And Celebrity Photos,” https://tinyurl.com/nohznmr, 2014.
[5] P. Mahajan, S. Setty, S. Lee, A. Clement, L. Alvisi, M. Dahlin, and M. Walfish, “Depot: Cloud Storage with Minimal Trust,” ACM Trans. Comput. Syst., vol. 29, no. 4, pp. 1–38, Dec. 2011.
[6] C. Gentry, S. Halevi, and N. P. Smart, “Homomorphic evaluation of the AES circuit,” in CRYPTO’12. Springer, 2012, pp. 850–867.
[7] P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes,” in EUROCRYPT’99, 1999, pp. 223–238.
[8] L. Weng, L. Amsaleg, A. Morton, and S. Marchand-maillet, “A Privacy-Preserving Framework for Large-Scale Content-Based Information Retrieval,” TIFS, vol. 10, no. 1, pp. 152–167, 2015.
[9] J. Z. Wang, J. Li, and G. Wiederhold, “SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 9, pp. 947–963, 2001.
[10] H. Muller, W. M ¨ uller, D. M. Squire, S. Marchand-Maillet, and T. Pun, “Performance evaluation in content-based image retrieval: overview ¨ and proposals,” Pattern Recognit. Lett., vol. 22, no. 5, pp. 593–601, 2001.