ISSN: 2321-7782 (Online)

Volume 3, Issue 1, January 2015

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Confidentiality Preserving Image Search using distance preserving randomization technique

Shinde Mangesh[1]
computer science
SCSCOE, Dhangwadi, Bhor,Dist- pune
Pune, India

Sontakke Pankaj[2]
Computer science
SCSCOE, Dhangwadi,Bhor,Dist- pune
Pune, India

Suryawanshi Vishal[3]
computer science
SCSCOE, Dhangwadi,Bhor,Dist- pune
Pune, India

Koli Swati[4]
Computer science
SCSCOE, Dhangwadi,Bhor,Dist- pune
Pune, India

Prof. Bhagwan D. Thorat[5]
Asst. Professor
Computer Department
SCSCOE,Dhangwadi,Bhor Dist-pune
Pune,India

***Abstract:** We propose a fully homomorphic encryption scheme – i.e., a scheme that allows one to evaluate circuits over encrypted data without being able to decrypt. Our solution comes in three steps. First, we provide a general result – that, to construct an encryption scheme that permits evaluation of arbitrary circuits, it suffices to construct an encryption scheme that can evaluate (slightly augmented versions of) its own decryption circuit; we call a scheme that can evaluate its (augmented) decryption circuit bootstrappable*

From few past years have seen increasing popularity of storing and managing personal multimedia data using online services. Preserving confidentiality of online personal data while offering efficient functionalities thus becomes an important and pressing research issue. In this paper, we study the problem of content-based search of image data archived online while preserving content confidentiality.

I. INTRODUCTION

Secure management of personal data stored online is an increasingly important issue, which demands a balance between data confidentiality and availability. Technologies that can enable secure online data management are going to be critically important. Existing privacy protection for online personal data focuses on access control and secure data transmission to ensure that the data can be securely transmitted to the server and unauthorized access to the data is prohibited. In order tp provide services to the user server decrypt received data and work on plain text, this in turn makes the user's private information vulnerable to untrustworthy service providers and malicious intruders. If data is stored online as plain text then administrator can view the stored data. Encrypted data makes difficult for server to provide services to the user. In order to handle this problem, it is both desirable and necessary to develop technologies for information retrieval over encrypted databases that can protect users' privacy without sacrificing the usability and accessibility of the information.

To build a secure CBIR system, both images and features should be protected. For a feature based retrieval system, images can be encrypted separately using cryptographic ciphers or image encryption algorithms. This paper focuses on the Problem of image feature protection which allows the computation of similarity measures among encrypted features, so that secure CBIR can be achieved.

II. Page Layout

Approach:

When using a standard web-based image search engine, one is likely to find garbage output even in the first couple of pages of results. Nevertheless, for simple objects, it can be conjectured that most of the top results will contain the object of interest. Our system exploits this consistency, and attempts to find the object that appears in most of these images. Preserving confidentiality of online personal data while offering efficient functionalities thus becomes an important and pressing research issue. In this paper, we study the problem of content-based search of image data archived online while preserving content confidentiality.

III. Existing System

Homomorphic encryption: Homomorphic encryption is a form of encryption which allows specific types of computations to be carried out on cipher text and generate an encrypted result which, when decrypted, matches the result of operations performed on the plaintext. This is a desirable feature in modern communication system architectures. Homomorphic encryption would allow the chaining together of different services without exposing the data to each of those services, for example a chain of different services from different companies could 1) calculate the tax 2) the currency exchange rate 3) shipping, on a transaction without exposing the unencrypted data to each of those services Homomorphic encryption schemes are malleable by design. The homomorphic property of various cryptosystems can be used to create secure voting systems, collision-resistant hash functions, and private information retrieval schemes and enable widespread use of cloud computing by ensuring the confidentiality of processed data.

There are several efficient, partially homomorphic cryptosystems, and a number of fully homomorphic, but less efficient cryptosystems. Although a cryptosystem which is unintentionally homomorphic can be subject to attacks on this basis, if treated carefully homomorphism can also be used to perform computations securely

Related Work:

Prior work in the area of information retrieval in the encrypted domain focused on text documents.

In explored Boolean search to identify whether a query term is present in an encrypted text document. In proposed a framework for rank-ordered search over encrypted text documents, so that documents can be returned in the order of their relevance to the query term. Secure text retrieval techniques can also be applied to keyword based search of image data. As we know keyword search relies on having accurate text description of the content already available, and its search scope is confined to the existing keyword set.

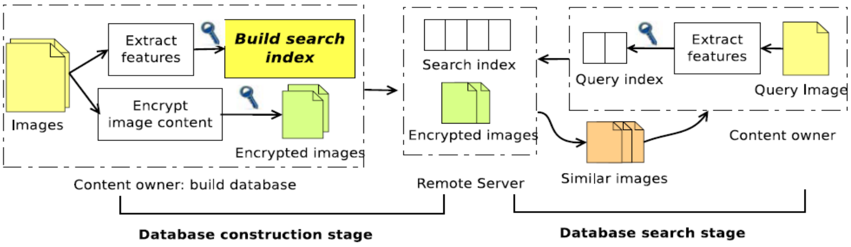
In contrast, content-based search over an encrypted image database provides more flexibility, whereby sample images are presented as queries and documents with similar visual content in the database are identified.

Current work in the area of secure computation for privacy protection addressed related but different problems under various application settings. considered privacy preserving range query over geospatial coordinates using a k-dimensional tree. Using such technique to image retrieval is difficult because features used for content based image retrieval are high dimensional vectors and kd-tree is known to be inefficient in high dimensional spaces. proposed secure k-NN computation that can determine which of two encrypted database entries has a smaller distance to the query, while keeping the actual distance value secret. We studied privacy preserving face recognition, where one party verifies the existence of a given face image in a database hosted on another party's servers. The two parties want to keep their own data secret from each other. Additive homomorphic encryption schemes are used to allow similarity computation in the encrypted domain.

IV. METHODOLOGY

Randomization Techniques for Visual Features and Search Indexes:

High computational and communication complexity involved in using homomorphic encryption for the task of rank-ordered image search, proposed to address the problem from a practical perspective and ask what can be done now as efficient solutions for this kind of practical applications.



A system model for the confidentiality preserving search scenario is shown in Fig. 1, where the left part depicts the database construction stage and the right part depicts the search stage. There are two entities in this model: a user who owns the private image collections, and a server who stores the encrypted data and performs retrieval based on a given encrypted query. During database construction, the user encrypts the images using standard ciphers and protects visual features or search indexes. After encryption, the user sends the encrypted images and protected features/indexes to the server for storage. During search, the user sends the similarly protected visual feature or search index of the query image to the server, who performs search using only the protected features or indexes. Finally, a list of encrypted images are ranked by their similarity to the query and returned to the user

Distance Preserving Randomization of Visual Features:

This feature introduce randomization is to scramble the content of visual features but approximately preserve the distance between features after randomization. Three types of feature randomization schemes are proposed namely, bit-plane randomization, random projection, and the randomized unary encoding.

1. Bit-Plane Randomization :-

The idea of bit-plane randomization is that feature vectors with small distances are likely to have similar patterns among their most significant bit-planes (MSB). Feature vectors are represented using binary form. To clamber the feature vector, each bit-plane is XORed with a binary random vector and then randomly permuted.

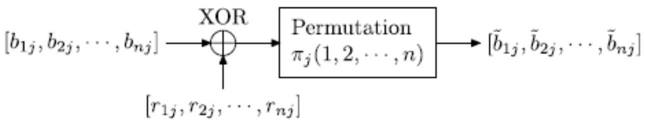
2. Random Projection:

This (RP) method is based on the idea that close points in high dimensional space will remain close with high probability after projection onto a low dimensional space and has been used as a building block for developing efficient search techniques for large databases. Random projection can be used to obfuscate the original values of the feature vectors while approximately preserving their distance.

Mathematical Modeling:

1. Bit-Plane Randomization:-

Given a feature vector **f=[f1,….,fn] € Rn** each component *fi* is represented in its binary form as [*bi*1; : : : ; *bil*]T , where *bi*1 is the first MSB, *bil* is the least significant bit (LSB), and *l* is the total number of bit-planes. The *j*th bit-plane of ***f*** is composed of the *j*th MSB of the *n* feature components, denoted as [*b*1*j*; *b*2*j*; : : : ; *bnj*]. To scramble the feature vector, each bit-plane is XOR ed with a binary random vector and then randomly permuted. The XOR pattern and permutation for the same *j*th bit-plane is the same so that Hamming distance between corresponding bit-planes is exactly preserved. The randomization of the *j*th bit-plane is illustrated in Fig. where [*r*1*j*; *r*2*j*; : : : ; *rnj*] is the binary random vector used for XOR.


$[b_{1j}, b_{2j}, \cdots, b_{nj}]$
XOR
Permutation
$\pi_j(1, 2, \cdots, n)$
$[\tilde{b}_{1j}, \tilde{b}_{2j}, \cdots, \tilde{b}_{nj}]$
$[r_{1j}, r_{2j}, \cdots, r_{nj}]$

All the randomized bit-planes are reassembled to form the randomized feature vector E(**f**)=[*f*1; : : : ; *fn*]. The distance between two randomized feature vectors E(**f**) and E(**q**) is computed using a weighted sum of Hamming distances between their individual bit-planes

$$d_{\mathcal{E}}(\mathcal{E}(\mathbf{f}), \mathcal{E}(\mathbf{g})) = \sum_{i=1}^{n} \sum_{j=1}^{l} |\tilde{b}_{ij}^{(\mathbf{f})} - \tilde{b}_{ij}^{(\mathbf{g})}| \times w(j)$$

$$d_{\mathcal{E}}(\mathcal{E}(\mathbf{f}), \mathcal{E}(\mathbf{q})) = \sum_{i=1}^{n} \sum_{j=1}^{l} |\tilde{b}_{ij}^{(\mathbf{f})} - \tilde{b}_{ij}^{(\mathbf{q})}| \times 2^{-j}.$$

dε(ε(f), ε(q)) = distance between two randomized feature vector.

This distance metric between randomized features is an upper bound on the *L*1 distance between the original features:

$$d_{\mathcal{E}}(\mathcal{E}(\mathbf{f}), \mathcal{E}(\mathbf{g})) \geq \sum_{i=1}^{n} \left| \sum_{j=1}^{l} (b_{ij}^{(\mathbf{f})} - b_{ij}^{(\mathbf{g})}) \times 2^{-j} \right| = \|\mathbf{f} - \mathbf{g}\|_1$$

$$\mathcal{E}(\mathbf{f}) = \mathbf{R} \cdot \mathbf{f} = [\mathbf{r}_1 \cdot \mathbf{f}, \ldots, \mathbf{r}_m \cdot \mathbf{f}] \in \mathbb{R}^m,$$

$$\begin{aligned} & d_{\mathcal{E}}(\mathcal{E}(\mathbf{f}), \mathcal{E}(\mathbf{g})) \\ & \quad = \sum_{i=1}^{m} |\mathbf{r}_i \cdot \mathbf{f} - \mathbf{r}_i \cdot \mathbf{g}| = \sum_{i=1}^{m} \|\mathbf{r}_i\|_2 \cdot \|\mathbf{f} - \mathbf{g}\|_2 \cdot |\cos(\theta_i)| \\ & \quad = \|\mathbf{f} - \mathbf{g}\|_2 \cdot \sum_{i=1}^{m} \|\mathbf{r}_i\|_2 \cdot |\cos(\theta_i)| \approx c \cdot \|\mathbf{f} - \mathbf{g}\|_2 \end{aligned}$$

Here is an independent and identically distributed random variable representing the angle between the vector **f** - **q** and the random vector **r***i*. By the law of large numbers, const and const. Thus, the distance between randomized features is proportional to the *L*2 distance between the original feature vectors with high probability

Andomized Unary Encoding :

$$\begin{aligned} \mathcal{U}(\mathbf{f}) &= [\mathcal{U}(f_1), \mathcal{U}(f_2), \ldots, \mathcal{U}(f_n)], \\ \mathcal{U}(f_i) &= \underbrace{11\cdots11}_{f_i}\underbrace{00\cdots00}_{M-f_i}. \end{aligned}$$

Distance Preserving Randomization Of Search Indexes:

For high dimensional vectors, comparing every pair of such vectors when doing a search is computationally prohibitive for a large database.

A. Secure Inverted Index

Inverted index is a widely used indexing structure in text document retrieval, where each keyword has an associated inverted index listing the documents that contain this keyword and the number of occurrences of this word in each of these documents. Only documents that appear in the query word's inverted index need to be considered during retrieval. By utilizing the visual words representation of images inverted index can be constructed for image documents and facilitates efficient search and retrieval over large image databases.

In order to generate inverted index, a vocabulary tree is first created, where each node in the tree denotes a representative feature vector and each leaf node represents a visual word. Such a vocabulary tree can be constructed using hierarchical k-means clustering on a set of training features. Given the vocabulary tree, each visual feature of an image will be treated as a word *R* and assigned to the closest visual word in the vocabulary tree.

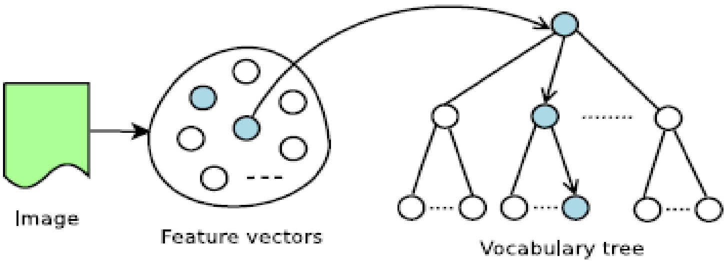


FIGURE 2. Inverted index generation by content owner.

Image ID		I2	...	I_{N_i}
Word frequency	w_1	w_2	...	w_{N_i}

FIGURE 3. Data structure of inverted index.

The similarity of two images is then measured by the Jaccard similarity between E(*Q*) and E(*D*):

$$\mathrm{Sim}(Q, D) \triangleq \frac{|\mathcal{E}(Q) \cap \mathcal{E}(D)|}{|\mathcal{E}(Q) \cup \mathcal{E}(D)|} \triangleq \frac{\sum_{i=1}^{V} \min(\mathcal{E}(Q_i), \mathcal{E}(D_i))}{\sum_{i=1}^{V} \max(\mathcal{E}(Q_i), \mathcal{E}(D_i))}.$$

As the order information used in min(_; _)

$$\mathcal{A}(Q_{MH}) = \{X_1^1, \ldots, X_1^{Q_1}, \ldots, X_N^1, \ldots, X_N^{Q_N}\},$$
$$\mathcal{A}(D_{MH}) = \{X_1^1, \ldots, X_1^{D_1}, \ldots, X_N^1, \ldots, X_N^{D_N}\}.$$

Here, *Xj*

i is a unique element indexed by *i* and *j*. The min-Hash values generated from A(*Q*) and A(*D*) are essentially elements randomly selected from the two sets, and they satisfy and max(_; _) is preserved by the order preserving encryption, the Jaccard similarity computed from the encrypted indexes reflects the similarity of the plaintext indexes.

$$\Pr[m(\mathcal{A}(Q_{MH}), f) = m(\mathcal{A}(D_{MH}), f)] = \mathrm{Sim}(Q_{MH}, D_{MH}) = \frac{\sum_{i=1}^{N} \min(Q_i, D_i)}{\sum_{i=1}^{N} \max(Q_i, D_i)}.$$

References

1. D. Song, D. Wagner, and A. Perrig, ‘‘Practical techniques for searches in encrypted data,’’ in *Proc. IEEE Symp. Res. Sec. Privacy*, Feb. 2000, pp. 44–55.
2. R. Brinkman, J. M. Doumen, and W. Jonker, ‘‘Using secret sharing for searching in encrypted data,’’ in *Proc. Workshop Secure Data Manag. Connected World*, 2004, pp. 18–27.

3. D. Boneh, G. Crescenzo, R. Ostrovsky, and G. Persiano, ‘‘Public-key encryption with keyword search,’’ in *Proc. Eur.*, 2004, pp. 506–522.

4. A. Swaminathan, Y. Mao, G.-M. Su, H. Gou, A. L. Varna, S. He, *et al.*,‘‘Confidentiality preserving rank-ordered search,’’ in *Proc. ACM Workshop Storage, Sec.*, *Survivability*, 2007, pp. 7–12.

5. Z. Erkin, A. Piva, S. Katzenbeisser, R. L. Lagendijk, J. Shokrollahi, G. Neven, *et al.*, ‘‘Protection and retrieval of encrypted multimedia content: When cryptography meets signal processing,’’ *EURASIP J. Inf. Sec.*, vol. 7, no. 2, pp. 1–20, 2007.

6. R. Datta, D. Joshi, J. Li, and J. Z. Wang, ‘‘Image retrieval: Ideas, influences, and trends of the new age,’’ *ACM Comput. Surveys*, vol. 40, no. 2, pp. 1–5, 2008.

7. J. Shashank, P. Kowshik, K. Srinathan, and C. Jawahar, ‘‘Private content based image retrieval,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

8. M.-L. Yiu, G. Ghinita, C. S. Jensen, and P. Kalnis, ‘‘Outsourcing search services on private spatial data,’’ in *Proc. IEEE 25th Int. Conf. Data Eng.*, Apr. 2009, pp. 1140–1143.

9. W. K. Wong, D. W.-L. Cheung, B. Kao, and N. Mamoulis, ‘‘Secure kNN computation on encrypted databases,’’ in *Proc. 35th SIGMOD Int. Conf. Manag. Data*, 2009, pp. 139–152.

10. Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, and T. Toft, ‘‘Privacy-preserving face recognition,’’ *Privacy Preserving Tech- nol., LNCS*, vol. 5672, pp. 235–253, Aug. 2009.

11. A.-R. Sadeghi, T. Schneider, and I. Wehrenberg, ‘‘Efficient privacy- preserving face recognition,’’ in *Proc. 12th Int. Conf. Inf. Sec. Cryptol.*, 2009, pp. 229–244.

12. M. Osadchy, B. Pinkas, A. Jarrous, and B. Moskovich, ‘‘Scifi—A sys- tem for secure face identification,’’ in *Proc. IEEE Symp. Sec. Privacy*, May 2010, pp. 239–254.