Privacy-preserving In the Cloud

[1]Ms.Chaitra G, [2]Mrs.Adarsh M S, [3]Mrs Kumaraswamy V S, [4]Prof Venkatesh
Dept. of Computer Science Engineering
Alva's Institute of Engineering and Technology
Mijar-574225, Mangalore, India

***Abstract*: Millions of private images are generated in various digital devices every day. The consequent massive computational workload makes people turn to cloud computing platforms for their economical computation resources. Meanwhile, the privacy concerns over the sensitive information contained in outsourced image data arise in public. In fact, once uploaded to cloud, the security and privacy of the image content can only presume upon the reliability of the cloud service providers. Lack of assuring security and privacy guarantees becomes the main barrier to further deployment of cloud-based image processing systems. In this paper studies the design targets and technical challenges lie in constructing cloud-based privacy-preserving image processing system. We explore various image processing tasks, including image feature detection, digital watermarking, content-based image search. The state-of-the-art techniques, including secure multiparty computation, and homomorphic encryption are investigated. A detailed taxonomy of the problem statement and the corresponding solutions is provided.**

Key Words—Cloud Computing; Image Process; Security; Cryptography; Privacy Preservation.

I. INTRODUCTION

Motivated by the rapid growth of image processing and data mining techniques, more and more image processing-based applications are deployed in various end-user's devices. For example, content-based image search, digital watermark verification and so on. The consequent massive image processing tasks bring enormous computation overhead to data owners. To solve this problem, more and more users are outsourcing the "expensive" tasks to cloud computing platforms. In such cloud computing platform, Cloud Service Provider (CSP) offers a pay-per-use business model, which enables individual user to use robust computation power in cloud while saving time and cost on setting up corresponding infrastructures. In fact, not only individual or small business data owners refer to, Internet giants like Microsoft and Yahoo are also attracted by the benefits brought by cloud computing and authorize some services to third-party cloud computing platforms. For example, several types of data searching tasks in Microsoft Bing have been outsourced to Wolfram.

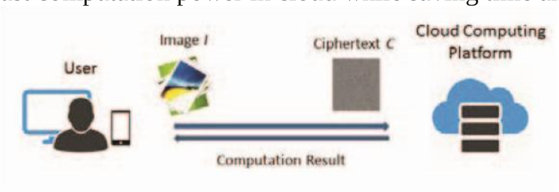

Fig. 1. System Model

However, the participation of a third-party cloud computing platform also increases the vulnerability of private data, e.g., potential data breach and lost. Under current cloud architecture, the content of outsourced image data will inevitably be leaked to CSPs. In this case, the leaked content might be sensitive information like data owner's personal identity, home address, or even financial records. Moreover, even we assume CSPs are completely honest and could be trusted to have data owners' private information, such privacy leakage still happens. In fact, cloud server is usually considered as a low-qualified locker rather than a strong bank deposit box. Comparing with traditional network server, the cloud computing platform suffers from more security threats. For instance, a severe vulnerability in cloud server is the sharing of computing resources: flaws in System Virtual Machine (SVM) software are frequently discovered and exploited to attack cloud servers in recent years. Nevertheless, the private data leakage in public cloud happens very often due to the improper configuration and maintenance by CSPs. In a nutshell, the privacy concern over the outsourced data has become the main barrier to the further development of cloud computing platforms.

In recent years, the secure image data processing is a rapidly growing research field and has attracted attention from both academia and industry. In practice, many fancy image processing applications require the computational power beyond the limit of mobile device for example, 3D structure econstruction needs massive computational power for image feature detection and matching. In this area, the main research direction lies in the detection of image features over ciphertext domain. Many encryption techniques are applied or adjusted to protect image data privacy while enabling visual feature extractions. In, a global image feature detection mechanism for color histogram-based descriptors detection is proposed. We utilize a Somewhat Homomorphic Encryption (SHE) scheme to enable the computation of diverse color descriptors in MPEG-7 standard over ciphertext domain. These features are further utilized as basic building blocks for services such as image matching and semantic tag generation. In a local feature detection mechanism for Scaler Invariant Feature Transform (SIFT) is proposed, which utilizes the Paillier encryption scheme to enable the computation of SIFT features over ciphertext domain. Moreover, in the authors analyze different scaling ratios by adjusting fixed point numbers in the proposed scheme. However, all these works suffer from the high computational complexity brought by homomorphic operations, especially for those who perform relatively complicated algorithms like SIFT. In the authors solve this problem by utilizing a multi-server structure to enable SIFT algorithm over encrypted data. In addition, another thriving research direction is the secure digital watermarking. It enables outsourcing the time-consuming tasks of generating digital watermark without compromising the privacy of the image content. Two types of approaches have been proposed: the asymmetric watermarking and the zero-knowledge watermark detection. However, most existing works still suffer from the high computational complexity on both user and cloud side.

Moreover, as an orthogonal research direction, the secure image retrieval mechanism is proposed in which enables applications such as location-based detection. It offers flexible approaches to manage private image datasets online. In the features extracted from images are encrypted in a distance-preserving scheme to enable direct comparisons for similarity evaluation. In the current image search indices are encrypted while achieving searching functionalities with efficiency.

However, in a practical privacy preserving computation scenario, all the existing works are very difficult to achieve the security requirements and practical efficiency performance at the same time. Here, the contributions of this paper are summarized as follows: We introduce and formulate diverse image processing tasks in a gener.al image computation outsourcing model, including image feature detection, digital watermarking, content-based image search. The state-of-the-art techniques, including secure multiparty computation and homomorphic encryption are discussed. A detailed taxonomy of the problem statement and the corresponding solutions are provided.

II. PRIVACY PROTECTION IN IMAGE DATA PROCESSING

A. System Model and Workflow

1) System Model:

As shown in Fig. 1, the proposed system consists of two main entities: Cloud Computing Platform (CCP) and user. User is a data owner who holds massive image data and intends to outsource the image processing tasks to the CCP. Under this setting, a user utilizes the CCP as a complementary resource for its limited computational power and outsources complicated image processing tasks to the CCP. Meanwhile, users also need to protect the privacy of data. For example, hospitals are under the obligation to protect the patient's records like medical images and profiles. In this case, to protect a user's privacy, he/she has to encrypt the image data before outsourcing to the CCP. Meanwhile, the entity CCP is composed by a set of cloud servers. It is assumed to be honest-but-curious. It can only access the encrypted image data uploaded by users and perform the corresponding image processing algorithms over ciphertext domain. After that, the CCP returns the requested results in the form of ciphertext back to a user. Finally, a user can use its private key to decrypt the returned results. Throughout the whole process, the CCP should not have any access to the content or results of the user outsourced image computation tasks in plaintext domain.

2) Workflow:

The proposed system consists of two main phases as follows:

- *Data Pre-processing*: In Data Pre-processing phase, for image *I*, a user prepares ciphertext *C* through encoding process *Encode* (*I*) and sends *C* to the CCP, where computation tasks over the encrypted image *C*. Such encoding algorithm should be lightweight and support as many image processing algorithms as possible. Hence, user only needs to encode its image data once, and the majority of computation workload is taken by CCP.
- *Encrypted Image Evaluation*: After receiving the encrypted image data, CCP performs image processing algorithms over the ciphertext domain to get the corresponding encrypted results. Meanwhile, the private information of uploaded image data should be protected against CCP. (After that, the user can decrypt and get image processing results in plaintext.)

Note that in this system architecture, users can get the maximum flexibility and scalability to perform massive image processing tasks. In fact, if a user has to perform a part of an image processing task and then upload the encrypted intermediates to CCP, the user's flexibility will be limited. Under this circumstance, a user will have to compute and encrypt different intermediates for various image processing tasks respectively. Nevertheless, even a minor parameter change in processing algorithms will force the user to compute and encrypt the whole image dataset over again.

B. Design Targets

After building the system model and defining the workflow, we formulate the design targets in constructing a privacy-preserving image processing mechanism on cloud: The first design target should be functionality, which requires the proposed system to perform image processing algorithms and generate corresponding results correctly. The second design target should be security, which requires the proposed system to protect image contents' confidentiality against the CCP while performing the processing algorithms on ciphertext domain. The last design target should be

efficiency, which requires the computation complexity on both the user and the CCP side and the communication complexity between them to be practical. These three design targets are equally important. However, if we must set a priority, the most important target should be security. After all, the sensitive information leakage can result in severe loss. Here, we use image feature detection algorithms as a set of case studies to analyse above three design targets:

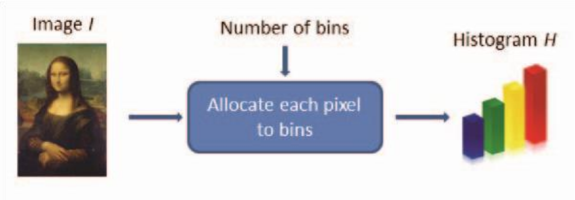

Fig. 2. RGB Histogram

1) Functionality:

As we discussed in Section I, image feature detection algorithms can be divided into two main categories: global feature detection, e.g., RGB histogram, Color Layout Descriptor (CLD), Color Structure Descriptor (CSD) and so on, and local feature detection, e.g., Here we use the functionality of RGB histogram as an illustrative example for global feature detection algorithms. In color feature detection algorithms, histogram descriptor is the most basic descriptor and building blocks for advanced feature descriptors. Based on color histogram, we can compute a series of prevalent color descriptors, including CSD, CLD. As shown in Fig. 2, the computation algorithm of color histogram in plaintext is very simple. However, if we intend to perform this algorithm over ciphertext domain, the functionality requirement makes it very difficult to be realized by simple encryption schemes: We need to enable the comparison between ciphertext and plaintext to correctly distribute every pixel value into the color histogram. Intuitively, this functionality requirement seems to be contradictive to the design target of security, or the confidentiality of encryption image data. If ciphertexts are comparable to plaintexts, the adversary can easily deduce all the values of encrypted pixels and get the sensitive information contained in an image. However, after carefully analyzing the functionality requirement of histogram algorithm, we can find that the exact required functionality is not the result of comparison between ciphertext and plaintext. The actually required functionality is the corresponding comparison result in ciphertext domain. Based on this observation, we initialize a somewhat homomorphic encryption scheme to ful-fill the corresponding functionality requirements and develop a privacy-preserving image global feature detection algorithm based on it. The corresponding experimental details are described in the paper.

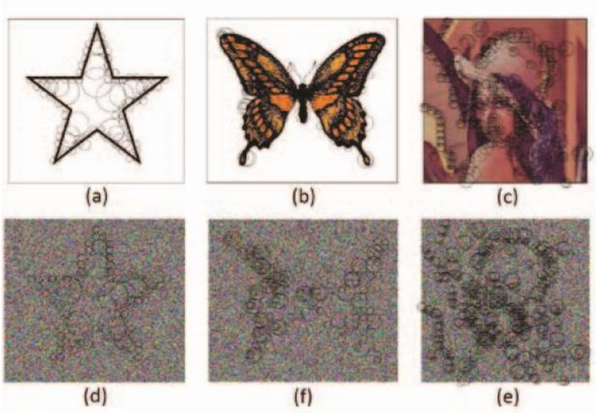

Fig. 3. Illustrative experimental result of SIFT Feature Descriptor: Figures (a-c) are the results in plaintext domain. Figures (d-f) are the corresponding results in ciphertext domain.

2) Security:

Recall that in the system model described above, we assume that CCP to be honest-but-curious. It means that the CCP will follow the procedures in the protocol and correctly perform the feature detection algorithm over ciphertext domain to protect its credits for the commercial benefits. However, it is still easy for an adversary, e.g., curious cloud engineer, to deduce the sensitive information contained in the image data through monitoring the data flow in ciphertext domain. Specifically, this kind of attack is especially hard to be defended in performing local feature detection algorithms. Here, we use SIFT as an illustrative example from local feature detection algorithms: As a local feature detection algorithm, SIFT algorithm first needs to detect the location of interesting points in an image. After that, it characterizes interesting points neighbour pixels by generating the corresponding feature descriptors around it. In Fig. 3, circles with different sizes represent different local feature descriptors, whose centers are the location of interesting points. In the process of secure image processing, since the CCP needs to generate those local descriptors, it will inevitably deduce the location of those interesting point in the image. However, from Fig.3, we can find that an eavesdropper on ciphertext domain data flow can easily get rough shapes of objects in the image. Through analysis, we can discover that this problem is similar to the pixel value comparison problem we met in color histogram algorithm. Is this problem can be solved by following the same methodology? More specifically, is it possible to solve this problem through encrypting the location of pixels and enabling the detection of interesting points on ciphertext domain? Unfortunately, this methodology can only convert the problem from the contradiction between security and functionality to the contradiction between functionality and efficiency. In fact, based on the complexity analysis of the corresponding method, it is easy to find that additional computational complexity for hiding pixel positions equals to the computational complexity of brute force attack against the encryption scheme. To achieve the functionality requirements, it seems to be impossible for the proposed system to provide a practical efficiency performance under the traditional definition of data confidentiality in cryptography. To solve this problem, authors introduce a multi-server structure-based mechanism to achieve a balance among the functionality, security, and efficiency simultaneously.

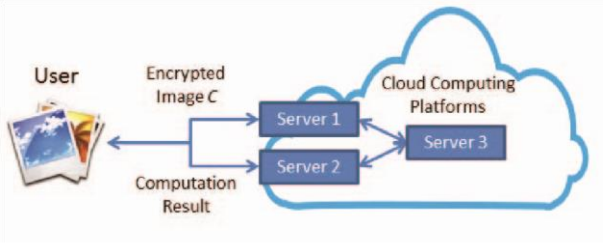

Fig. 4. Utilizing SMC techniques by introducing additional cloud computing platform

3) *Efficiency:*

In the complexity analysis of secure cloud computing, we need to analyse the efficiency of the proposed mechanism in three aspects: The computational complexity on the user and the CCP side, and the communication complexity between these two parties. In practice, to achieve the flexibility on user's side and scalability on the CCP's side, most existing designs only allocate necessary procedures like encryption and decryption tasks to the user. Consequently, complicated functionalities are required in the corresponding mechanisms. As a result, it leads to more complicated encryption algorithms to be applied and finally overloads user's computational complexity. With respect to a few homomorphic encryption algorithms, the corresponding encryption and decryption computation complexity is even larger than the computation complexity of performing the image processing algorithm. In this case, the practice of the corresponding mechanism is neglected. Hence, not only we need to develop the encryption scheme that can provide enough homomorphic operations required in the image processing algorithm, but also have to carefully balance the computation and communication complexity to ensure the feasibility of the proposed design

III. SMC BASED IMAGE PROCESSING

Usually, Secure Multiparty Communication protocol is considered as a general solution to any function computations. However, since its enormous computation and communication complexity, it is not widely implemented in practice. Nevertheless, its advantage on compatibility and simplicity of the SMC algorithm makes it play a very important role in secure cloud computing mechanism designs. Among many SMC techniques, the Secure Two-party Computation is often utilized as a building block in constructing the system with techniques like homomorphic encryption scheme.

1) *SMC based Secure Image Feature Detection*:

In image feature detection algorithms, the functionality requirements like comparison, factorial, and trigonometric operations exist in many complicated image feature detection algorithms. However, these operations over ciphertext domain require tens to hundreds of iterations of homomorphic additions and multiplications operations. Hence, it seems to be impractical to use only homomorphic encryption-based techniques to realize all those functionalities. To solve this problem, one possible methodology is adjusting the system architecture of cloud computing platform to utilize SMC techniques. As shown in Fig. 4, user can easily realize homomorphic additions and ciphertext comparisons through introducing additional cloud servers. For example, a simple implementation of one-time-pad encryption scheme from SMC protocols that splits one plaintext into two ciphertexts enables homomorphic additions. This design methodology can be generalized to utilize various arithmetic encryption methods from SMC protocols. Moreover, the utilization of SMC technique also provides an additional perspective to balance the communication complexity and computation complexity. As shown in Fig. 4, in the framework of SMC technique, it is effective to alleviate computation complexity on user side by introducing additional communication complexity between user and cloud servers, e.g., through uploading ciphertexts to two cloud servers.

2) *SMC based Secure Image Digital Watermarking:*

Another popular implementation of SMC technique is in secure image watermark detection algorithms. A digital watermark is a signal permanently embedded in digital data, i.e., audio, pictures, video. This signal is hidden in the host data in such a way that it is inseparable from the data and so that it is resistant to many operations not degrading the host digital data. The watermark can be detected or extracted later through computing operations in order to make assertions about the data. Various secure digital watermarking system models are proposed. In existing works, cloud is usually can be utilized to perform tasks like watermark generation, detection, and matching. Among them, one typical model in watermark detection is shown in Fig. 5.

To construct a secure watermark detection mechanism, most existing solutions leverage SMC techniques.

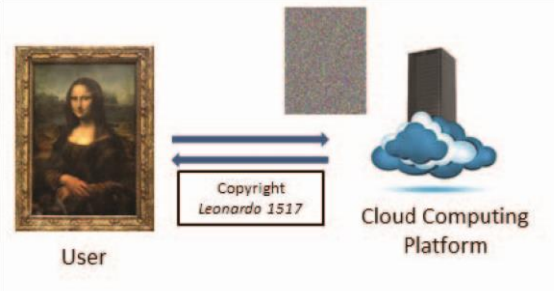

Fig. 5. Workflow of Secure Digital Watermarking Detection in Cloud

image feature detection applications, the proposed secure sharing scheme divides user's data into multiple pieces and uploads them to different cloud servers, making it difficult to derive the whole information from any one cloud. Another work focuses on the efficiency of the video data watermarking, which splits the original video and uses Hadoop distributed computing system for the different requirements to realize watermarks embedding. a framework for message privacy-preserving copy detection and watermark identification based on the signs of the Discrete Cosine Transform (DCT) coefficients is proposed. The architecture allows for searching in encrypted data and places the computational overhead on cloud server. The low frequency DCT coefficients are utilized to generate a dual set of keys to encrypt the source image, and a robust hash for the digital watermarking queries.

Moreover, by utilizing SMC technique, some secure watermarking tasks performed on CCP side have shown close performance as performed in the plaintext domain. Using random matrix transformation, which can be considered as a two-dimensional extension of the one-time-pad technique, an efficient privacy-preserving watermarking detection mechanism is proposed. In a nutshell, using SMC techniques, privacy-preserving watermarking processing mechanisms are secure and efficient under certain conditions.

However, though utilizing SMC techniques can effectively reduce computational complexity comparing with using HE techniques. The inherent feature of SMC techniques requires that the data owner remains to be online when cloud performs most operations. Hence, it only applies to limited types of applications in practice.

IV. CONCLUSION AND FUTURE WORK

In this paper the problem of privacy-preserving image processing on cloud computing platform, which could enable any fancy image processing-based applications on devices with limited computation power. For example, a variety of instant image processing apps on the lens, watch or other personal devices. Comparing with other outsourced computation tasks, image processing algorithms are relatively complicated and have high computation complexity. To solve the problem, we start with building system model and formulating design targets. After that, the state-of-the-art techniques are introduced, including homomorphic encryption, secure multiparty computation and so on. We also present several case studies for different techniques and analyse their merits and drawbacks. Through the analysis, we find that the balance among design targets: functionality, security, and efficiency make it difficult to solve the problem by applying only one technique. The integration of different techniques other than traditional cryptography tools is the most promising research direction in this area. In addition, considering the prevalence of JPEG compression among age data, which can be considered. Privacy-preserving decompression of JPEG file as a special case of privacy-preserving DCT computation is also a promising research direction in this area.

REFERENCES

[1] C.-Y. Hsu, et al. Homomorphic encryption-based secure SIFT for privacy-preserving feature extraction. In Proc. of SPIE, 2011.

[2] M. Naehrig, et al. Can homomorphic encryption be practical? In Proc. of CCSW, 2011.

[3] College Professors about cloud.