Ciphertext-Only Attack on an Image Homomorphic Encryption Scheme with Small Ciphertext Expansion

Yunyu Li, Jiantao Zhou, and Yuanman Li
Faculty of Science and Technology, University of Macau
{mb35468, jtzhou, mb25510} @umac.mo

ABSTRACT

The paper "*An Efficient Image Homomorphic Encryption Scheme with Small Ciphertext Expansion*" (*In Proc.* ACM MM'13, pp.803-812) presented a novel image homomorphic encryption approach achieving significant reduction of the ciphertext expansion. In the current work, we study the security of this cryptosystem under a ciphertext-only attack (COA). We show that our proposed COA is effective in generating a sketch of great fidelity of the original image. Experimental results are provided to verify the validity of the proposed attack strategy.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Information Systems Applications- Miscellaneous; K.6.5 [**Computing Milieux**]: Metrics—*Management of Computing and Information Systems–Security and Protection*

Keywords

Image homomorphic encryption, ciphertext expansion, ciphertext-only attack

1. INTRODUCTION

Signal processing over encrypted domain (SPED) has been receiving increasing attention in recent years, primarily driven by the various privacy-preserving applications and the wide adoption of cloud computing platforms [7, 2, 8]. Among the encryption solutions enabling SPED, homomorphic encryption is arguably the most popular one, as it provides a generic framework of performing basic algebraic operations over the encrypted domain [11, 10, 5, 6]. The existing homomorphic cryptosystems can be roughly classified into two categories: partially homomorphic schemes [5, 10] and fully homomorphic ones [6]. The partially homomorphic schemes support encrypted domain operations that correspond to the addition *or* multiplication in the plaintext domain, while the latter type allows both the addition *and* multiplication.

However, one of the major obstacles that precludes the widespread adoption of homomorphic encryption in practice is the huge expansion of the ciphertext. For instance, when the Paillier cryptosystem, with modulus being 1024 bits, is used to encrypt 8-bit image, the resulting encrypted file is 256 times larger than the original image [10]. For fully homomorphic cryptosystem, the ciphertext expansion is even much more prohibitive[6]. To deal with the ciphertext expansion problem, Trobcoso-Pastoriza proposed a packing scheme, in which several messages are packed as a word and encrypted together [12]. This scheme was later extended in [4] with generalized packing basis. However, such packing makes it impossible to process each message differently, and causes many operations, e.g., encrypted-domain DFT [3] and DWT [13], infeasible without interactive protocol. Recently, Zheng and Huang suggested a very promising approach for reducing the ciphertext expansion by indexing a sequence of ciphertexts produced by the scaled-down histogram of the image [14]. It was claimed that significant reduction of the ciphertext expansion can be achieved, while not affecting the security of the image encryption system.

In this work, we evaluate the security of the image homomorphic encryption scheme in [14] under a ciphertext-only attack (COA). We design a state set in which each element represents an encrypted (quantized) pixel, and model the image local smoothness property through a state transition matrix describing the correlation between different states. The statistical analysis of this transition matrix allows us to determine, with high accuracy, which locations share similar (even identical) pixel values, and which ones more likely lie in the edge regions. We also design a pixel value assignment strategy to further preserve image local smoothness and enlarge the image contrast, enabling us to obtain a sketch version of the original image of high quality. Experimental results are provided to verify the effectiveness of our proposed COA strategy.

The rest of this paper is organized as follows. Section 2 gives a concise introduction of the image homomorphic encryption scheme [14]. In Section 3, we present our proposed COA strategy. Experimental results are given in Section 4 to demonstrate the effectiveness of our method. We finally conclude in Section 5.

2. IMAGE HOMOMORPHIC ENCRYPTION SCHEME PROPOSED IN [14]

The idea of this image homomorphic encryption scheme is to first generate a pixel sequence $\mathbf{S} = (S_0, S_1, \cdots, S_{K-1})$ where K is smaller than the number of pixels in the input

MM'15, October 26–30, 2015, Brisbane, Australia.
© 2015 ACM. ISBN 978-1-4503-3459-4/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2733373.2806406.

image, and then produce a map with each element being the index of the original pixel value relative to $\mathbf{S}$. The homomorphically encrypted sequence $[[\mathbf{S}]]$ and the index map in the plaintext serve as the ciphertext. In the sequel, we use $[[x]]$ to represent the element-wise, homomorphically encrypted version of x.

More specifically, let $\mathbf{I} = \{I(x, y)\}$ be the image to be encrypted, where $0 \leq x, y \leq M - 1$. If not otherwise specified, we assume images to be 8-bit and Paillier as the homomorphic cryptosystem [10]. Let also $\mathbf{n} = (n_0, n_1, \cdots, n_{255})^T$ be the associated histogram vector, where n_i denotes the number of pixel value i in $\mathbf{I}$. The sequence $\mathbf{S}$ is designed in such a way that 1) $0 \leq S_j \leq 255, \forall j$, 2) the associated histogram vector $\mathbf{v} = (v_0, v_1, \cdots, v_{255})^T$ is a scaled-down version of $\mathbf{n}$, namely

$$v_i = \lceil \frac{n_i}{Q} \rceil \tag{1}$$

where $Q > 1$ is a scaling factor; and 3) the order of placing all S_j is *randomly* determined. The sequence $\mathbf{S}$ is encrypted element-wise into $[[\mathbf{S}]] = ([[S]]_0, [[S_1]]_1, \cdots, [[S]]_{K-1})$, using a probabilistic homomorphic encryption scheme (Paillier), in which secret key is involved. This also implies that $[[S]]_j \neq [[S]]_k$ if $j \neq k$.

To produce the index map, the sequence $[[\mathbf{S}]]$ is divided into a series of sub-sequences $[[\mathbf{S}]](i)$, for $0 \leq i \leq 255$, each of which corresponds to the original pixel value i, i.e.

$$[[\mathbf{S}]](i) = \Big\{ [[s]] \in [[\mathbf{S}]] \Big| \mathcal{D}([[s]]) = i \Big\} \tag{2}$$

where $\mathcal{D}(\cdot)$ denotes the homomorphic decryption function. The input image $\mathbf{I}$ is then partitioned into non-overlapping blocks, and the index map generation is carried out in a block by block fashion. For each pixel value $I(x, y)$, one component from the sub-sequence $[[\mathbf{S}]](I(x, y))$ is selected sequentially and its index in the whole sequence $[[\mathbf{S}]]$ is recorded in the index map. In the case that all the elements in $[[\mathbf{S}]](I(x, y))$ are consumed, a randomly permuted $[[\mathbf{S}]](I(x, y))$ is used. Following the notation of [14], we use $\mathbf{P}([[\mathbf{I}]], [[\mathbf{S}]])$ to denote the index map relative to $[[\mathbf{S}]]$.

Eventually, the ciphertext is represented by $\{[[\mathbf{S}]], \mathbf{P}([[\mathbf{I}]], [[\mathbf{S}]])\}$, which can be transmitted to a third party, e.g., cloud server for outsourced computation. Prior to homomorphic operations, an element-wise, homomorphically encrypted image $[[\mathbf{I}]]$ can be calculated based on $[[\mathbf{S}]]$ and $\mathbf{P}([[\mathbf{I}]], [[\mathbf{S}]])$ in a straightforward and cost-effective way (simple lookup table process).

A very desirable feature of this image homomorphic encryption scheme is the significantly reduced ciphertext size. Assuming the modulus used in Paillier is of 1024 bits, the ciphertext size contributed by the encrypted sequence $[[\mathbf{S}]]$ is $2048 \cdot K$ bits. For the index map $\mathbf{P}([[\mathbf{I}]], [[\mathbf{S}]])$, it can be readily coded with fixed-length coding and the size can be determined by $M^2 \cdot \lceil \log_2 K \rceil$. Therefore, the size of the encrypted image becomes

$$2048 \cdot K + M^2 \cdot \lceil \log_2 K \rceil \tag{3}$$

The ciphertext expansion factor can then be given by

$$\gamma = \frac{256 \cdot K}{M^2} + \frac{\lceil \log_2 K \rceil}{8} \tag{4}$$

Typical values of K are 1024, 2048, and 4096 (corresponding Q values are 256, 128, and 64) for 512×512 sized images, which can lead to significant reduction of the ciphertext expansion [14]. In contrast, in the traditional image homomorphic encryption scheme, the ciphertext expansion factor is fixed to be 256 when the modulus is of 1024 bits. It was also claimed that high level of security can still be achieved. For more details, please refer to [14].

3. CIPHERTEXT-ONLY ATTACK

When evaluating the security of the image homomorphic encryption scheme in [14], we assume that the Paillier cryptosystem is secure. Following Kerckhoff's principle, the strength of a cryptosystem depends only on the key and, in particular, the security does not depend on keeping the encryption algorithm secret. This principle implies that the attacker knows the protocols and the overall system in which the cryptosystem is used, while only does not know the secret key. According to the information that is available to the attacker, the attacks can be classified into several types.

- A ciphertext-only attack (COA) is one where the attacker tries to deduce the secret key or the plaintext by only observing the ciphertext.
- A known-plaintext attack (KPA) is one where the attacker has a quantity of plaintext and the corresponding ciphertext
- A chosen-plaintext attack (CPA) is one where the attacker chooses plaintext and is then given the corresponding ciphertext
- A chosen-ciphertext attack (CCA) is one where the attacker can have access to the decoder, and thus can select ciphertext and obtain the corresponding plaintext

Though COA is the least effective attack type, it is of practical significance, as in many applications the ciphertext is the only information available to the attacker. As to be demonstrated shortly, the image encryption scheme [14] is vulnerable even under COA.

More specifically, the attacker can access the ciphertext composed of $[[\mathbf{S}]]$ and $\mathbf{P}([[\mathbf{I}]], [[\mathbf{S}]])$. Since Paillier is assumed to be secure, we cannot infer any useful information from $[[\mathbf{S}]]$. Our COA strategy instead exploits the information leaked from the index map $\mathbf{P}([[\mathbf{I}]], [[\mathbf{S}]])$, based on the following key observations: 1) identical encrypted pixel values can be observed in the encrypted image $[[\mathbf{I}]]$, and they must correspond to the same original pixel value; and 2) pixels that are close spatially should have similar or even identical original values, though the encrypted ones may differ significantly. The latter observation is due to the local smoothness property inherent to natural images.

With $[[\mathbf{S}]]$ and $\mathbf{P}([[\mathbf{I}]], [[\mathbf{S}]])$, the attacker can generate the encrypted image $[[\mathbf{I}]]$, as the cloud server does. To model the local smoothness of image signal in encrypted domain (as only encrypted signal is available), we define a state set

$$\mathcal{S} = \Big\{ [[S]]_0, [[S]]_1, \cdots, [[S]]_{K-1} \Big\} \tag{5}$$

where each distinct encrypted pixel value is regarded as a state. To further integrate the image local smoothness prior,

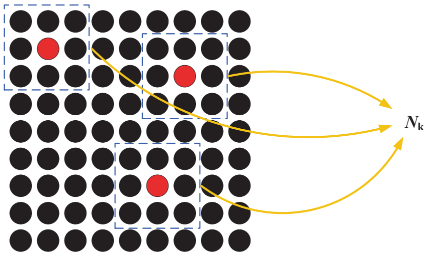

Figure 1: The process of generating $\mathcal{N}_k$. Here, red dots denote $[[S]]_k$.

we propose to quantize the state set $\mathcal{S}$ according to the spatial distribution of encrypted pixels. To this end, we localize $[[S]]_k$ in the encrypted image $[[\mathbf{I}]]$ (may have multiple locations, as $K < M^2$). We then construct a 3×3 sized patch centered by each localized $[[S]]_k$, and use $\mathcal{N}_k$ to denote the collection of patches ($[[S]]_k$ excluded), as demonstrated in Fig. 1. We randomly select τ elements from $\mathcal{N}_k$ and eliminate them from $\mathcal{S}$. A typical setting is $\tau = 3$. This is equivalent to quantize these τ elements into $[[S]]_k$. This process is repeated for all the remaining, unprocessed $[[S]]_k$, which eventually results in a quantized set

$$\mathcal{S}' = \left\{[[S]]'_0, [[S]]'_1, \cdots, [[S]]'_{K'-1}\right\} \tag{6}$$

where $K' = \lceil K/(\tau + 1)\rceil$. A state transition matrix $\mathbf{T}$ can then be defined to describe the relationship between different states of $\mathcal{S}'$

$$\mathbf{T} = \begin{bmatrix} T(0,0) & T(0,1) & \cdots & T(0,K'-1) \\ T(1,0) & T(1,1) & \cdots & T(1,K'-1) \\ \vdots & \vdots & \ddots & \vdots \\ T(K'-1,0) & T(K'-1,1) & \cdots & T(K'-1,K'-1) \end{bmatrix} \tag{7}$$

where the state transition probability $T(k,l)$ from state $[[S]]'_k$ to $[[S]]'_l$ can be calculated as follows. We first localize all $[[S]]'_k$ in the encrypted image $[[\mathbf{I}]]$, and form the set $\mathcal{N}'_k$ in a very similar way as Fig. 1 illustrates. $T(k,l)$ is then computed by

$$T(k,l) = \frac{\#[[S]]'_l \in \mathcal{N}'_k}{|\mathcal{N}'_k|} \tag{8}$$

where $|\cdot|$ denotes the cardinality of the corresponding set.

The state transition matrix $\mathbf{T}$ obtained completely in encrypted domain conveys important information about the original image. When $T(k,l)$ is large, it means that $[[S]]'_l$ appears in the neighboring region of $[[S]]'_k$ with high probability. Due to the local smoothness of image signal, the original pixel value S'_l and S'_k very likely are similar or even identical. On the other hand, when $T(k,l)$ is relatively small (but nonzero), e.g., median value among all $T(k,l)$, $0 \leq l \leq K'-1$, it implies that $[[S]]'_l$ appears in the neighboring region of $[[S]]'_k$ with non-negligibly small (not very high neither) probability. In this case, the original pixels S'_l and S'_k may reside in different sides of an edge, and should be remarkably different. Hence, if we assign identical values for those $[[S]]'_k$ and $[[S]]'_l$ with $T(k,l)$ being large, and sufficiently different values for those $[[S]]'_k$ and $[[S]]'_l$ with $T(k,l)$ being relatively small, then we can expect a sketch version of the original image. This is because natural images are primarily formed by smooth objects separated by boundaries and edges.

More specifically, we can employ the following **Algorithm 1** to reconstruct a sketch of the original image from the available encrypted image $[[\mathbf{I}]]$ and the pre-calculated state transition matrix $\mathbf{T}$. Here, the med$(\cdot)$ and AHE$(\cdot)$ denote the median function and the adaptive histogram equalization (AHE) function, respectively. The purpose of applying the AHE is to enhance the contrast of the resulting sketch image, making it visually more pleasing. It should be noted that we make the step size of updating the pixel values in the sketch image rather small ($m \leftarrow (m+1) \mod 256$). This helps better preserve the image local smoothness characteristics, as we process the pixels in a raster-scan order and adjacent locations (except high-activity regions) should have similar pixel values.

Algorithm 1 Obtain a sketch image from the ciphertext and the state transition matrix

```
Input: $\mathbf{P}([[\mathbf{I}]], [[\mathbf{S}]])$, $[[\mathbf{S}]]'$, and $\mathbf{T}$
Output: The sketch image $\hat{\mathbf{I}} = \{\hat{I}_{x,y}\}$
Generate the encrypted image $[[\mathbf{I}]]$ according to $\mathbf{P}([[\mathbf{I}]], [[\mathbf{S}]])$ and $[[\mathbf{S}]]'$.
Initialization: $\hat{\mathbf{I}} \leftarrow [[\mathbf{I}]]$, $m \leftarrow 0$
for $x = 0$ to $M-1$ do
  for $y = 0$ to $M-1$ do
    if $\hat{I}_{x,y}$ has been filled in then
      continue
    end if
    Determine the state index $k$ associated with $[[\mathbf{I}(x,y)]]$
    Computer two quantities
        $l_{\max} = \arg\max_l T(k,l)$    (9)
        $l_{\text{med}} = \text{med}\left\{T(k,l)\right\}_{l=0}^{K'-1}$    (10)
    Replace all $[[S]]'_{l_{\max}}$ in $\hat{\mathbf{I}}$ with pixel value $m$
    Replace all $[[S]]'_{l_{\text{med}}}$ in $\hat{\mathbf{I}}$ with pixel value $255 - m$
    Update $m \leftarrow (m+1) \mod 256$
  end for
end for
$\hat{\mathbf{I}} \leftarrow \text{AHE}(\hat{\mathbf{I}})$
```

4. EXPERIMENTAL RESULTS

In this section, we experimentally evaluate the performance of the proposed attack strategy using two groups of test images. The first group consists of three images, two of which are from the MNIST image database [9], and the other one is a medical image. These three images are relatively simple with large portion of homogeneous regions. In Fig. 2, we show the quality of the sketch images, where the first column lists the original images, and the last three columns give the corresponding sketch images obtained by running the **Algorithm 1** when the scaling factor $Q = 2, 4, 6$, respectively. It can be seen that even when Q is rather small, e.g., $Q = 4$, the sketch images are of high quality. For instance, the digits 1 and 4 are largely recognizable. With the increasing of Q values, higher quality of the sketch images

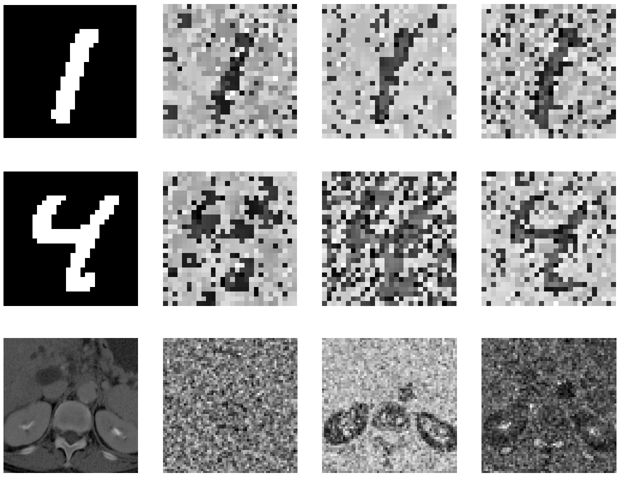

Figure 2: Reconstructed sketch images for Digit 1 (first row), Digit 4 (second row), and a medical image Liver (third row) when $Q = 2, 4, 6$, respectively.

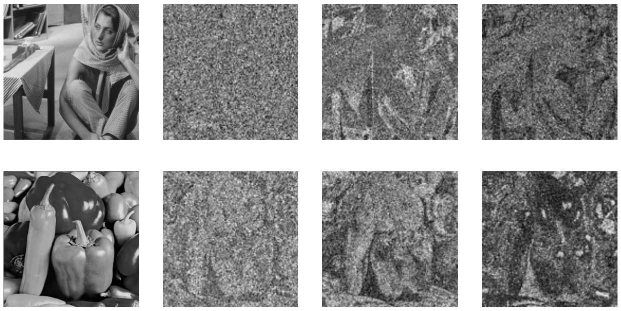

Figure 3: Reconstructed sketch images for Barbara (first row) and Peppers (second row) when $Q = 4, 6, 8$, respectively.

can be retained. This is due to the fact that more repeated encrypted pixel values would appear in $[[\mathbf{I}]]$, making the task of estimating the sketch images more effective.

The second test group is composed of two widely-used natural images selected from the CVG-UGR gray level image database [1]. The sketch images illustrated in the last three columns of Fig. 3 are retained when the scaling factor $Q = 4, 6, 8$, respectively. It can be observed that when $Q \geq 6$, rich amount of semantic information of the original image can be recovered. Further, the quality of the sketch images improves when Q becomes larger.

As a secure image encryption method should be successful in protecting *any* images, the above experimental results on a limited test set are sufficient to demonstrate the security vulnerability of the image encryption scheme in [14].

5. CONCLUSIONS

In this paper, we have analyzed the security problem of a recently published image homomorphic encryption scheme with small ciphertext expansion. We have shown that even under ciphertext-only attack (COA), the least effective attack type, the attacker can still recover a rich amount of semantic information from the encrypted images. Experimental results have been provided to validate our findings.

6. ACKNOWLEDGMENTS

This work was supported in part by the Macau Science and Technology Development Fund under grants FDCT/009/2013/A1 and FDCT/046/2014/A1, in part by the Research Committee at University of Macau under grants MRG007/ZJT/2015/FST, MRG021/ZJT/2013/FST, MYRG2014-00031-FST and MYRG2015-00056-FST, and in part by the National Science Foundation of China under grant 61402547.

7. REFERENCES

[1] CVG UGR image database. http://decsai.ugr.es/cvg/dbimagenes/.

[2] C. Aguilar Melchor, S. Fau, C. Fontaine, G. Gogniat, and R. Sirdey. Recent advances in homomorphic encryption: A possible future for signal processing in the encrypted domain. *IEEE Sig. Proc. Mag.*, 30(2):108–117, 2013.

[3] T. Bianchi, A. Piva, and M. Barni. On the implementation of the discrete fourier transform in the encrypted domain. *IEEE Trans. Inf. Forensics Security*, 4(1):86–97, 2009.

[4] T. Bianchi, A. Piva, and M. Barni. Composite signal representation for fast and storage-efficient processing of encrypted signals. *IEEE Trans. Inf. Forensics Security*, 5(1):180–187, 2010.

[5] T. ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Trans. on Inf. Theory*, 31(4):469–472, 1985.

[6] C. Gentry. Fully homomorphic encryption using ideal lattices. In *Proc. 41st Annu. ACM Symp. Theory Comput. (STOC'09)*, pages 169–178, 2009.

[7] C.-Y. Hsu, C.-S. Lu, and S.-C. Pei. Image feature extraction in encrypted domain with privacy-preserving sift. *IEEE Trans. on Image Proc.*, 21(11):4593–4607, 2012.

[8] R. L. Lagendijk, Z. Erkin, and M. Barni. Encrypted signal processing for privacy protection: Conveying the utility of homomorphic encryption and multiparty computation. *IEEE Sig. Proc. Mag.*, 30(1):82–105, 2013.

[9] Y. LeCun. MNIST handwritten digit database. Available on the web at http://www.research.att.com/yann/ocr/mnist/.

[10] P. Paillier. Public-key cryptosystems based on composite degree residuosity classes. In *Advances in Cryptology (EUROCRYPT'99)*, pages 223–238. Springer, 1999.

[11] R. L. Rivest, L. Adleman, and M. L. Dertouzos. On data banks and privacy homomorphisms. *Foundations of Secure Computation*, 4(11):169–180, 1978.

[12] J. R. Troncoso-Pastoriza, S. Katzenbeisser, M. Celik, and A. Lemma. A secure multidimensional point inclusion protocol. In *Proc. 9th ACM Workshop on Multimedia and Security*, pages 109–120. ACM, 2007.

[13] P. Zheng and J. Huang. Discrete wavelet transform and data expansion reduction in homomorphic entrypted domain. *IEEE Trans. Image Proc*, 22(6):2455–2468, 2013.

[14] P. Zheng and J. Huang. An efficient image homomorphic encryption scheme with small ciphertext expansion. In *Proc. of the 21st ACM Int. Conf. on Multimedia (MM'13)*, pages 803–812. ACM, 2013.