ON ENCRYPTION AND AUTHENTICATION OF THE DC DCT COEFFICIENT

Li Weng and Bart Preneel*
Department of Electrical Engineering, Katholieke Universiteit Leuven, 3001 Heverlee, Belgium

Keywords: Encryption, Authentication, DC coefficient, DCT.

Abstract: When encryption and authentication techniques are applied to image or video data, sometimes it is advantageous to limit the operation to the DC DCT coefficient of each 8×8 block in a picture. In this work, the performance of such an approach is evaluated. This problem is considered as an image quality problem, and the metric structural similarity is used to show that by authenticating the DC coefficient, about 60% of the information can be guaranteed; by encrypting the DC coefficient, about 80% of the information can be hindered.

1 INTRODUCTION

With the increasing awareness of security in multimedia applications, conventional encryption and authentication techniques are also being adopted to protect multimedia data. The former guarantees that the data is not accessible to any unauthorized party; the latter guarantees data integrity, and sometimes also verifies the source (Stallings, 2006). In spite of their different purposes, when confronted with the huge amount of multimedia data, especially for video, they both suffer from significant computation complexity. For example, current High Definition TV systems have data rates up to 10s of Mbps for a single channel. Such high data rates can already make significant burden for video servers. With the extra requirement of encryption and authentication, the overall design of a video transmission system becomes even more complicated. Therefore, reducing processing overhead is an essential concern for these applications.

Among various video or image encryption and authentication solutions, those that operate in the *discrete cosine transform* (DCT) domain seem to show practical interest, e.g., (Lin and Chang, 2001; Zeng and Lei, 2003), because currently the most popular video or image standards, such as MPEG-1/2/4 and JPEG, all store visual information in the DCT domain. If encryption or authentication is directly applied to data in the DCT domain without going back to the spatial domain, computation for the inverse DCT can be saved. Therefore, the DCT domain is favored by both encryption and authentication.

However, it is not always feasible to process all the data in the DCT domain. In some scenarios, only part of the DCT coefficients are encrypted or authenticated, e.g., (Weng et al., 2006; Sun et al., 2006). This might be due to low processing power or real-time constraints, as well as other particular requirements. As a result, there is a typical approach consisting of encrypting or authenticating *only* the DC coefficient. For example, an MPEG authentication scheme which is robust to transcoding was proposed in (Sun et al., 2006), where the DC coefficient is authenticated; an MPEG encryption algorithm with multiple security levels was proposed (Li et al., 2007; Weng et al., 2006) and the first level is DC coefficient en-

*This work was supported in part by the Concerted Research Action (GOA) Ambiorics 2005/11 of the Flemish Government and by the IAP Programme P6/26 BCRYPT of the Belgian State (Belgian Science Policy). The first author was supported by the IBBT-VIN project, which was co-funded by IBBT (Interdisciplinary Institute for BroadBand Technology), a research institute founded by the Flemish Government in 2004, and the involved companies and institutions (Philips, IPGlobalnet, Vitalsys, Landsbond onafhankelijke ziekenfondsen, UZ-Gent). https://projects.ibbt.be/vin

Weng L. and Preneel B. (2007).
ON ENCRYPTION AND AUTHENTICATION OF THE DC DCT COEFFICIENT.
In *Proceedings of the Second International Conference on Signal Processing and Multimedia Applications*, pages 365-369
DOI: 10.5220/0002137503650369

cryption. Empirically, it seems that the DC coefficient is a good candidate for encryption and authentication, and generally leads to a good compromise between computation overhead and security. However, to the best of our knowledge, there are neither theoretical nor experimental results to quantify *how much* information can be guaranteed or hindered by authenticating or encrypting the DC coefficient. Without such results, it is difficult for designers to decide whether to adopt similar approaches. Motivated by this fact, we would like to investigate the above problem in a quantitative way and measure the actual performance of DC coefficient encryption and authentication.

In this work, the evaluation of encryption and authentication performance is considered as an image quality evaluation problem. An image quality metric called "structural similarity" (Wang et al., 2004) is adopted to measure the amount of information hindered by partial encryption or guaranteed by partial authentication. From various experiments applied to an image database of about 420 JPEG images, we show that by authenticating DC coefficients, about 60% of the information can be guaranteed; by encrypting DC coefficients, about 80% of the information can be hindered. Although these are not absolute results, they might give insights into the problem.

The rest of the work is organized as follows: Section 2 first introduces our approach to evaluate encryption and authentication performance by structural similarity; Subsection 2.1 gives the background of partial image authentication and results of various experiments in terms of structural similarity; Subsection 2.2 is a similar approach to partial image encryption. Section 3 concludes our work.

2 BENCHMARKING DC COEFFICIENT ENCRYPTION AND AUTHENTICATION

Although encryption and authentication are quite different, a uniform approach can be used to evaluate their performance for visual data. Because they both consider how much information is preserved after the operation, one could think of them as an *image quality* problem. Assuming that the original image has full quality, for partial encryption, we measure the quality of the encrypted version compared to the original one; for partial authentication, we measure the quality of the authenticated part compared to the original one. Therefore, the problem is converted into the choice of a proper image quality metric.

The most widely used image quality metrics are the peak signal-to-noise ratio (PSNR) and the mean square error (MSE). They are simple and efficient for general purposes. However, "they are not very well matched to perceived visual quality" (Wang et al., 2004), because they only concentrate on the amount of errors, but not the perceived information. For these metrics, image quality measure is quite different from the amount of information expressed through the image. For example, Figure 1 shows a gray-scale Lena image (a) and several distorted versions: (b) subtracting a constant from all pixel values and setting negative results to zero; (c) applying an averaging filter; (d) JPEG compression. The PSNRs of the distorted versions, compared to the original one, are given below the images. Although the distorted versions look rather similar to the original one, the PSNRs are quite low. One can also note that the PSNR of (b) is lower than the one of (c), while (b) actually has better perceptual quality. Therefore, it might not be appropriate to use simple metrics that measure the error visibility, such as PSNR and MSE; instead, we need a metric which indeed measures image similarity.

We find that the image quality metric *structural similarity* (SSIM) (Wang et al., 2004) fulfills the requirement. This metric compares luminance, contrast, and structure information between two gray-scale images of the same size and returns an average score between zero and one, with one meaning exactly the same and zero meaning completely different. It is defined as:

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma}\ , \quad (1)$$

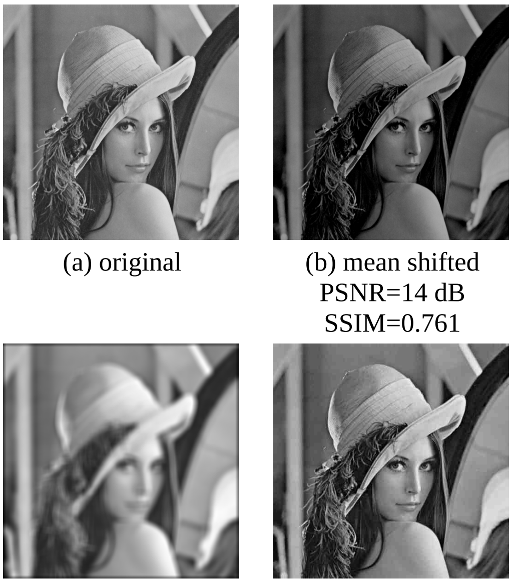
where x and y represent two test images; functions $l()$, $c()$, and $s()$ correspond to luminance, contrast, and structure similarity, respectively; α, β, and γ are weighting factors. In our experiments, we use its simplified form:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}\ , \quad (2)$$

where μ represents mean; σ represents (co)variance; C_1 and C_2 are numerical constants for stability (we use 6.5025 and 58.5225 as suggested). This metric satisfies the following conditions:

- Symmetry: $SSIM(x,y) = SSIM(y,x)$;
- Boundedness: $SSIM(x,y) \leq 1$;
- Unique maximum: $SSIM(x,y) = 1$ iff $x = y$.

These properties make it easy to interpret the meaning of an SSIM score. Due to space constraints, we skip elaboration on the details of this metric. For more information, one can refer to the paper (Wang et al., 2004). Although this metric does not cover all aspects of image similarity measure, it gives more reasonable



(a) original

(b) mean shifted
PSNR=14 dB
SSIM=0.761

(c) blurred
PSNR=22.3 dB
SSIM=0.654

(d) JPEG compressed
PSNR=25.7 dB
SSIM=0.855

Figure 1: Lena vs. distorted Lena.

results than MSE or PSNR in our scenario. Figure 1 also shows the SSIM scores for the distorted versions, which are reasonably high and more consistent with human perception.

In the following sections, we first introduce the background of DC coefficient authentication and encryption, then design experiments and measure the performance by SSIM. Because the purpose is to measure the importance of the DC coefficient, we limit our experiments to images. Since video can be considered as sequences of images, it is reasonable to assume that the experiment results also indicate the performance for video scenarios.

2.1 DC Coefficient Authentication

Authentication can achieve data integrity verification and/or source verification. In this work, only the former is considered. Due to space constraints, we skip the formal introduction to authentication techniques; we assume that once authentication is applied, the integrity of the involved data can be guaranteed.

Traditional authentication scenarios assume that the data that need to be protected never change (Stallings, 2006). However, multimedia data is sometimes subjected to transcoding during transmission. Typical transcoding techniques are: requantization, frame resizing, and frame dropping (Vetro et al., 2003). They all modify the original data, thus compromising the integrity check. If the authentication is applied to all DCT coefficients, it will fail at the verification stage while the content is still authentic. Therefore, a simple approach to circumvent this difficulty is to authenticate low-frequency DCT coefficients which are less likely to be affected by transcoding. For example, an MPEG authentication scheme which is robust to the above mentioned transcoding techniques was proposed in (Sun et al., 2006). The authors suggest to authenticate the DC coefficient of each 8×8 block and demonstrated acceptable results. However, they also mentioned that the selection of features for authentication is application-specific, without further motivating their choice. Our experiment studies the performance of their approach.

In order to evaluate the “importance” of the DC coefficient, we design the following experiment:

1. Divide a gray image into 8×8 blocks;
2. Perform DCT on each block;
3. Leave the DC coefficient and set all others to zero;
4. Perform inverse DCT and restore the image;
5. Compare the restored image with the original one by SSIM.

Figure 2(a) illustrates the effect of this procedure applied to the Lena image. Note that this is equivalent to replacing the pixel values in an 8×8 block by the mean of all pixels. This experiment is carried out for about 420 real-life JPEG images. They are divided into the following sets:

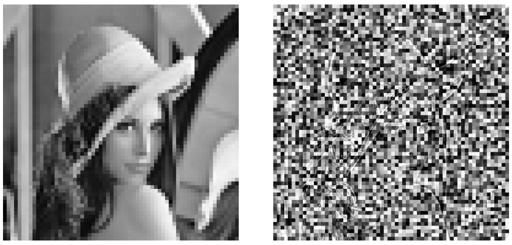
Type 1: architecture
Type 2: sculpture
Type 3: humanoid
Type 4: landscape
Type 5: objects
Type 6: vehicle

Each set contains around 70 images. They are converted to gray-scale and resized to three canonical sizes before the experiment:

Size 1: 640×480
Size 2: 1024×768
Size 3: 1600×1200

The average results from all sets are listed in Table 1. From the results one can see that the SSIM score is usually above **0.6**, and it increases with the size. That can be interpreted as, that more than 60% of the information of the original image is preserved in the DC coefficients. This is an interesting result, because the percentage of data representing DC coefficients in an image or video bitstream is obviously much less than that. Therefore, it shows that DC coefficient authentication is quite a cost-effective approach.

However, normally two different images might have an average SSIM score around 0.5. Compared to the average score of **0.662** from above experiment, it seems that the DC coefficient is not that significant. To give a fair comparison, we also measure the



(a) DC only
PSNR=23.3 dB
SSIM=0.626

(b) random DC
PSNR=9.1 dB
SSIM=0.179

Figure 2: Lena with: (a) only DC; (b) random DC.

Table 1: Average SSIM for DC coefficient authentication.

	Size 1	Size 2	Size 3	Average
Type 1	0.634	0.686	0.742	0.687
Type 2	0.577	0.609	0.655	0.614
Type 3	0.661	0.703	0.750	0.705
Type 4	0.641	0.655	0.689	0.662
Type 5	0.625	0.674	0.725	0.675
Type 6	0.583	0.623	0.673	0.626
Average	0.620	0.658	0.706	**0.662**

SSIM score between any two different images of the same size within each test set. The average results are around **0.3**, which is half of the score of the DC coefficients. This confirms that DC coefficients indeed contain significant information.

Since our experiments show that the DC coefficient carries so much information, it might be interesting to know how much information is carried by other low-frequency DCT coefficients. Therefore, we repeat the above experiment with some modification. Instead of leaving the DC coefficient only, we keep the first two DCT coefficients of each 8×8 block, and the rest of the experiment is the same. The average results are around **0.7**. It shows that about 70% of the information is preserved in the first two DCT coefficients, i.e., the gain is around 10% compared to authenticating DC coefficients alone. We noted that the gain is decreasing if we repeat this experiment for more DCT coefficients. Therefore, it might be less cost-effective to authenticate more DCT coefficients.

2.2 DC Coefficient Encryption

Encryption of multimedia data used to have a speed problem, especially for video. Therefore, partial encryption was advocated to alleviate the situation. Currently, although powerful computing seems to be less expensive, partial encryption is still interesting for some other purposes. For example, sometimes partial encryption is preferred, in order to make the encrypted bitstream *format-compliant* (Wen et al., 2002; Liu and Eskicioglu, 2003; Weng et al., 2006). Among various partial encryption approaches, encrypting the DC coefficient has special interest for video standards such as H.262 (MPEG-2 video) and H.263 (MPEG-4 simple profile). In H.262, the length of the DC coefficient is indicated beforehand; in H.263, the DC coefficient is coded as an 8-bit fixed-length field. Therefore, encrypting only the DC coefficient does not compromise the format. This can enable many interesting features, such as random frame access and simple statistics in the encrypted domain.

Table 2: Average SSIM after DC encryption.

	Size 1	Size 2	Size 3	Average
Type 1	0.163	0.146	0.128	0.146
Type 2	0.199	0.188	0.171	0.186
Type 3	0.173	0.160	0.145	0.159
Type 4	0.182	0.181	0.170	0.178
Type 5	0.183	0.164	0.146	0.164
Type 6	0.199	0.184	0.165	0.183
Average	0.183	0.171	0.154	**0.169**

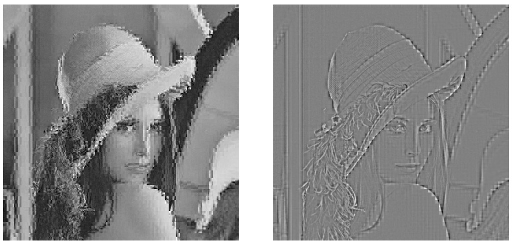
In order to simulate and measure the effect of encrypting DC coefficients, we make the following modification to the 3rd step of previous experiment:

3. Set the DC coefficient of each block to a random value (0-255) and leave all others;

Figure 2(b) illustrates the effect of this procedure when applied to the Lena image. It has almost become unrecognizable. Therefore this encryption approach seems to have good performance. This is confirmed by the new experiment after applying to the same image sets as in previous ones. The average SSIM scores are listed in Table 2. They are all below **0.2**, indicating that more than 80% of the information has been hindered. This is a very promising result.

As a comparison, we also test the performance of encrypting sign bits of DCT coefficients, which is another well-known partial encryption approach (Bhargava et al., 2004). We modify our experiment to randomly flip the sign of DCT coefficients except for the DC coefficient. Figure 3(a) shows the Lena image encrypted this way. The observed distortion is very limited. Therefore, this approach is not as effective as DC coefficient encryption. The results show an average score above **0.5**, which is consistent with our observation.

Nevertheless, note that encrypting DC coefficients or sign bits is sometimes vulnerable to error concealment attack (ECA) (Wen et al., 2002). For example, one can set the DC coefficient of each block to 128 to cancel some encryption effect. This is illustrated in Figure 3(b), where the encrypted Lena im-



(a) random DCT sign
PSNR=20.28 dB
SSIM=0.502

(b) error concealment
PSNR=14.70 dB
SSIM=0.750

Figure 3: Lena with: (a) random DCT sign; (b) ECA.

age in Figure 2(b) is subjected to ECA. After the attack the contour of Lena becomes quite visible and the SSIM score increases to **0.75**. Therefore, such kind of encryption schemes are only sufficient for scenarios where the security requirement is not very high, e.g., commercial video applications.

3 CONCLUSION AND DISCUSSION

When encryption or authentication is applied to image or video data in the DCT domain, the operation is sometimes limited to DC coefficients only. Although empirically plausible, we have evaluated the corresponding security gain in a quantitative way. We have considered this as an image quality problem and adopted the metric structural similarity: for authentication, we measure the SSIM between an original image and a quality-reduced version which only consists of DC coefficients; for encryption, we measure the SSIM between an original image and a distorted version whose DC coefficients are scrambled. After extensive experiments with about 420 real-life images of various types, we conclude that by authenticating DC coefficients, more than 60% of the information is guaranteed; and by encrypting DC coefficients, more than 80% of the information is hindered. These results coincide with empirical experience and on the other hand give more sound basis for DC-coefficient-based approaches. Our experiments are limited to images, but the results can also imply the performance when similar approaches are applied to video, due to the similarity between image and video coding.

Note however, that in this work, the way we measure the amount of information is not exact. Therefore our results can be used only as reference or guidelines. Our experiments are limited to gray-scale images, so the results might only explain the case of luminance part of visual data, which is usually the focus for many image operations. Extending on color images and video are interesting topics for future research.

REFERENCES

Bhargava, B., Shi, C., and Wang, S.-Y. (2004). MPEG video encryption algorithms. *Multimedia Tools Appl.*, 24(1):57–79.

Li, S., Chen, G., Cheung, A., Bhargava, B., and Lo, K.-T. (2007). On the design of perceptual MPEG-video encryption algorithms. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(2):214–223.

Lin, C.-Y. and Chang, S.-F. (2001). A robust image authentication method distinguishing JPEG compression from malicious manipulation. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(2):153–168.

Liu, X. and Eskicioglu, A. (2003). Selective encryption of multimedia content in distribution networks: Challenges and new directions. In *Proc. of IASTED Int. Conference on Communications, Internet and Information Technology*.

Stallings, W. (2006). *Cryptography and Network Security*. Prentice Hall, 4th edition.

Sun, Q., He, D., and Tian, Q. (2006). A secure and robust authentication scheme for video transcoding. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(10).

Vetro, A., Christopoulosa, C., and Sun, H. (2003). Video transcoding architectures and techniques: An overview. *IEEE Signal Processing Magazine*, pages 18–29.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4).

Wen, J., Severa, M., Zeng, W., Luttrell, M., and Jin, W. (2002). A format-compliant configurable encryption framework for access control of video. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(6):545–557.

Weng, L., Wouters, K., and Preneel, B. (2006). Extending the selective MPEG encryption algorithm PVEA. In *Proc. of IEEE Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Pasadena, USA.

Zeng, W. and Lei, S. (2003). Efficient frequency domain selective scrambling of digital video. *IEEE Transactions on Multimedia*, 5(1):118–129.