

© IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

ENCRYPTING ONLY AC COEFFICIENT SIGNS CONSIDERED HARMFUL

Heinz Hofbauer¹ • Andreas Unterweger¹ • Andreas Uhl¹

¹Multimedia Signal Processing and Security Lab, Department of Computer Sciences, University of Salzburg, Austria, {hhofbaue, aunterweg, uhl}@cosy.sbg.ac.at

Abstract

We show that the selective encryption of AC coefficient signs in DCT-based video formats is not suitable for use cases which require confidentiality, but can be used for some other application scenarios. By proposing a new assessment based on the method by Wu et al. we analyze the range of formats from the older JPEG standard to the more recent H.265/HEVC standard. We conclude that commonly used measures like exemplary PSNR values or key space calculations for theoretical brute force attacks are not sufficient to draw conclusions on the security of selective AC coefficient sign encryption.

Contents

| | |
|--|----------|
| 1 Introduction | 2 |
| 2 Methods for Analyzing Selective Encryption Approaches | 2 |
| 3 Tools and Data | 3 |
| 4 Analysis of AC Sign Encryption | 3 |
| 4.1 H.265/HEVC Analysis | 3 |
| 4.2 JPEG Analysis | 4 |
| 5 Conclusion | 5 |

1 Introduction

Selective (also: partial) encryption of multimedia data has many applications, ranging from privacy preservation in video surveillance [1, 2] to transparent encryption for TV broadcasts [3, 4]. Despite the different use cases, all applications share the requirement of content security, i.e., it must be hard for an attacker to obtain the unencrypted data from the encrypted data.

Due to the widespread use of DCT-based video compression standards, most of the literature on selective encryption focuses on approaches for these formats. Furthermore, since AC coefficient (ACC) signs are easily accessible and modifiable in most implementations (and often stored uncompressed or nearly uncompressed in the compressed bit stream), they are often used as the plain text for encryption. Examples for selective encryption approaches of this kind include approaches for MPEG-1/2 [5], MPEG-4 Part 2 [6], H.264/AVC [7, 8] (as well as its scalable extension [9]) and H.265/HEVC [10, 11].

Despite the widespread use of AC coefficient sign encryption (ACSE), security and attack analyses are very limited. A large percentage of papers do not include explicit security evaluations at all, while most of the others only show selected example images before and/or after simple attacks (e.g., [7, 10]), sometimes with PSNR values (e.g. [11]). Even fewer papers include calculations of theoretical key space sizes for brute-force attacks (e.g., [7]). We argue, however, that these analyses are not sufficient and additionally provide a false sense of security.

Typically, there is also no distinction between the goal of the encryption (examples for encryption goals are given in [12, 13]). Since selective encryption is weaker than traditional encryption (parts of the media stream are left unencrypted), it can target the following three application scenarios: *confidential encryption*, where the content must not be reconstructible; *sufficient encryption* with (very) low quality, where reconstructing a higher-than-target quality must not be possible; and *transparent encryption* with reduced quality. The latter is similar to sufficient encryption, but instead of a destructive low quality version an encrypted version with a given quality has to be constructed.



Figure 1: Examples of compressed (left), encrypted (middle) and attacked images (right): Frame 100 from *foreman* as JPEG with 75% (top) and H.265/HEVC with *randomaccess* GOP structure and QP 27 (bottom).

Traditional encryption is impossible with selective encryption, and from a simple set of example images (Fig. 1) it is obvious that confidential encryption can also not be achieved with ACSE. This leaves sufficient and transparent encryption as target applications.

We will demonstrate that ACSE in DCT-based video formats is, by itself, a problematic approach for sufficient and transparent encryption. We do so by describing an extended set of analyses which can be used for (future) selective encryption approaches that include ACSE.

This paper is structured as follows: In Section 2, we motivate and explain our new analyses. In Section 3, we describe the tools and data we used. Finally, in Section 4, we show our results before concluding the paper.

2 Methods for Analyzing Selective Encryption Approaches

The three following facts have to be considered when analyzing selective encryption approaches.

Firstly, encryption on a bit stream introduces an error in the decoded media. In the case of ACSE, the residual information which would ‘repair’ errors made by various forms of prediction is changed in such a way that not only the errors are not corrected, but amplified. In such a case, the expedient method for analyzing quality is to disregard the changes introduced by the encryption. This is usually done by a replacement attack which substitutes the encrypted part with elements which are statistically least likely to introduce errors. In the case of ACSE, this is setting the ACCs to zero, as shown in Fig. 1. The target for sufficient encryption of selective AC signs should therefore be equal to the attacked case, i.e., where the ACCs are set to zero.

Secondly, the key space is not an indicator of quality. Often, a combination of a few sample images and the fact that the key space is large are given as evidence for a sufficiently low quality, e.g., [7]. This is incorrect since a large key space only makes it more expensive for an attacker to break the encryption as a whole. In addition, for a successful attack, i.e., the reconstruction of a higher quality, color might not be necessary at all and consequently the key space may actually be significantly smaller than given.

Thirdly, key sensitivity has to be analyzed in terms of both, traditional key sensitivity and quality sensitivity. Traditionally, key sensitivity pertains to the chaotic nature of encryption, i.e., a small change (single bit flip) of the key should produce a vastly different cipher text. For visual media, one can simply change the key minimally and compare the quality of the encrypted decoded media. However, for selective encryption, the quality also has to be taken into consideration. This means that the choice of the key may change the output quality, which is undesired. Optimally, only the selection (which parts of the bit stream to encrypt) should impact the target quality. If higher-than-targeted quality can result from the choice of a poor key, the encryption scheme is flawed. Hence, the quality sensitivity of the key should be measured similarly to regular key sensitivity, i.e., by introducing small changes in the key. Instead of

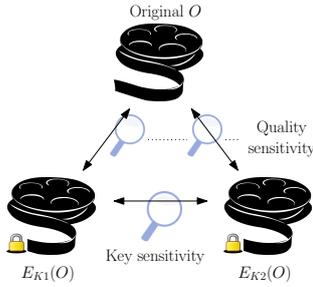


Figure 2: Illustration of the difference between key and quality sensitivity.

the quality between the decoded encrypted media, the average and standard deviation of the quality between the original and (a sufficiently high number of) decoded encrypted media should be given, see Fig. 2.

Key sensitivity is typically estimated by the number of pixel change rate (NPCR) and the unified average change intensity (UACI), together with a randomness test, see Wu et al. [14] for details for full image encryption. However, since ACSE is partial encryption where pixel change intensities depend on the magnitude of the residuals, the randomness test needs to be adjusted.

Given that all other assumptions from [14] hold, the possible amplitude of change is dependent on the residuals and thus dependent on the medium. Therefore, we will use the average UACI (U) between the original (O) and the attacked (decoded with zeroed ACCs, $Z(O)$) as the range for change in magnitude in two directions, positive or negative, $M = \min(L, U(O, Z(O)) \frac{2L}{100})$, where L is the maximum pixel value. The critical value for rejecting the randomness hypothesis then is $N_{\alpha}^* := \frac{1}{M+1} (M - \Phi^{-1}(\alpha) \sqrt{\frac{M}{WH}})$, where W and H are the width and height of the image and Φ^{-1} is the quantile function of the normal distribution.

3 Tools and Data

The information contained in ACCs (and their signs) highly depends on the compression format and its possibilities to eliminate redundancy. In JPEG [15], for example, ACCs are not predicted in any way, i.e., they are a representation of the image content. Conversely, in H.265/HEVC [16], numerous intra- and inter-prediction mechanisms are applied before the DCT, i.e., the ACCs are representations of the corresponding residual signals and are no longer directly related to the image content.

Formats like MPEG-4 Part 2 and H.264/AVC use fewer and/or less sophisticated prediction mechanisms than H.265/HEVC, but more than JPEG. The properties of their ACCs with regards to encryption are therefore in-between those of JPEG and H.265/HEVC. To cover a wide range of DCT-based formats, we focus on the two "extremes" JPEG and H.265/HEVC. We use the two reference software implementations *jpeg-8d* and *HM* to encode our test sequences *foreman* and *crew* (CIF).

For H.265/HEVC compression, we use the three default configurations – *intra* (I^* GOP structure), *lowdelay* ($IP\{4\}$), i.e.,

| | GOP structure | μ | σ | min. | max. |
|----------------|---------------------|-------|----------|-------|-------|
| <i>foreman</i> | <i>intra</i> | 98.00 | 0.13 | 97.36 | 98.40 |
| | <i>lowdelay</i> | 98.21 | 1.41 | 88.35 | 99.85 |
| | <i>randomaccess</i> | 98.32 | 0.56 | 95.44 | 99.44 |
| <i>crew</i> | <i>intra</i> | 98.68 | 0.07 | 98.45 | 98.91 |
| | <i>lowdelay</i> | 98.74 | 1.54 | 90.01 | 99.92 |
| | <i>randomaccess</i> | 98.66 | 0.37 | 96.90 | 99.45 |

Table 1: NPCR statistics over 4950 different key pairs.

IPPPP GOP structure) and *randomaccess* ($IB\{32\}$ GOP structure). We use quantization parameters (QPs) from 9 to 51 with a step size of 6 to achieve a wide range of quality levels.

For JPEG compression, we convert the video sequences to (bitmap) image sequences using *avconv*. Since JPEG is inherently limited to intra-picture coding, no GOP structure can be specified. We only vary the quality level from 0% to 100% in steps of 5%.

After compression, we encrypt both, the compressed JPEG and H.265/HEVC bit streams by toggling the ACCs signs based on a one-time pad. We attack the encrypted images by setting all ACCs to zero (not only the signs). Figure 1 depicts some examples of compressed, encrypted and attacked images.

4 Analysis of AC Sign Encryption

4.1 H.265/HEVC Analysis

For the key sensitivity and quality sensitivity tests we use the *foreman* and *crew* sequences with QP 21 which exhibited the highest amount of variation in terms of quality. We use 100 different keys, resulting in 4950 pairwise comparisons for the key sensitivity and 100 comparisons for the quality sensitivity.

Key Sensitivity — For *foreman/lowdelay*, the critical value for the NPCR randomness test is $N_{0.05}^*(L) = 99.1\%$ based on $U(L, Z(L)) = 22.79\%$. Assuming a normal distribution of the NPCR values with the $\mu = 98.21$ and $\sigma = 1.42$, the number of keys failing the randomness test is $\text{cdf}(x, \mu, \sigma) := \frac{1}{2}[1 + \text{erf}(\frac{x-\mu}{\sigma\sqrt{2}})]$. For *foreman/lowdelay* $\text{cdf}(99.1, 98.21, 1.42) = 73.6\%$ of the keys do not pass the randomness test.

The NPCR values of the *crew* and *foreman* sequences for the *lowdelay*, *randomaccess* and *intra* GOP structures are given in Table 1. The critical value and the results of the randomness test is given in Table 2.

None of the GOP-structure-sequence combinations pass the randomness test on average. The *intra* GOP structure is clearly the worst-performing in total due to the lack of temporal error propagation.

Quality Sensitivity — Quality sensitivity is analyzed similar to key sensitivity, but the quality is calculated between the original and the encrypted instead of between encrypted media. For *foreman/lowdelay* the PSNR value range is $[8.741, 15.396]$ with $\mu = 12.075$, $\sigma = 1.253$.

The changes in the signs of the ACCs due to encryption should introduce further errors rather than 'repair' prediction

| GOP structure | | $N_{0.05}^*$ | $U(O, \mathcal{Z}(O))$ | pass % |
|----------------|---------------------|--------------|------------------------|--------------|
| <i>foreman</i> | <i>intra</i> | 99.05 | 21.65 | $< 10^{-13}$ |
| | <i>lowdelay</i> | 99.10 | 22.79 | 26.40 |
| | <i>randomaccess</i> | 99.14 | 23.88 | 7.16 |
| <i>crew</i> | <i>intra</i> | 98.82 | 17.16 | 2.28 |
| | <i>lowdelay</i> | 99.09 | 22.53 | 41.01 |
| | <i>randomaccess</i> | 99.03 | 21.00 | 15.87 |

Table 2: Critical values for the randomness test and the percentage of keys which pass the test.

| GOP structure | | μ | σ | S^* | suff. % |
|----------------|---------------------|--------|----------|--------|--------------|
| <i>foreman</i> | <i>intra</i> | 11.244 | 0.118 | 12.346 | 100.00 |
| | <i>lowdelay</i> | 12.075 | 1.253 | 11.406 | 29.67 |
| | <i>randomaccess</i> | 11.129 | 0.664 | 11.210 | 54.85 |
| <i>crew</i> | <i>intra</i> | 12.747 | 0.165 | 15.241 | 100.00 |
| | <i>lowdelay</i> | 10.777 | 3.499 | 13.038 | 76.490 |
| | <i>randomaccess</i> | 12.322 | 0.828 | 13.742 | 95.68 |

Table 3: PSNR statistics, the threshold and the percentage of keys which result in sufficient quality.

errors. It stands to reason that the quality of an encryption based on ACCs should be worse than removing (zeroing) the ACCs altogether. Consequently, the quality of the medium with zeroed ACCs is used as a baseline for low quality. For *foreman/lowdelay* the PSNR is $S^* = 11.406$. Assuming a normal distribution with stable μ and σ , via cdf, we see that only 29.67% of the keys results in a sufficiently low quality (suff. %).

The results for this sufficiency test for other GOP structures for the *foreman* sequence as well as for the *crew* sequence are given in Table 3. Clearly, the GOP structure has a significant influence on the impact of the selective encryption and should not be disregarded.

In those cases where the encryption method achieves sufficient encryption (or any stable target quality), the method automatically allows for transparent encryption. As shown in [4], the target quality of ACSE can be steered in a monotonic fashion towards higher quality ranges by reducing the number of encrypted signs.

In the above cases, the *intra* GOP structure is very well suited for sufficient and transparent encryption. The *lowdelay* and *randomaccess* GOP structures, however, are not suited for sufficient encryption since the percentage of keys resulting in a quality low enough to serve as a starting point is small. For transparent encryption, the high variance in starting quality makes it difficult to consistently prevent a high quality preview. As such, it should only be used for low quality previews. This is obviously true for the *lowdelay* GOP structure (σ values). The standard deviation for the *randomaccess* GOP structure is significantly lower, but still too wide, especially since the starting points are most likely already in the transparent range of quality, i.e., higher than sufficient.

When considering key sensitivity as well as quality sensitivity, none of the GOP structures are well suited. The *lowdelay* GOP structure is not suited based on quality and the *intra* and

| Quality | | μ | σ | min. | max. |
|----------------|----|--------|----------|--------|--------|
| <i>foreman</i> | 25 | 68.931 | 0.626 | 66.262 | 71.169 |
| | 50 | 82.618 | 0.430 | 81.154 | 84.245 |
| | 75 | 89.249 | 0.254 | 88.147 | 90.057 |
| <i>crew</i> | 25 | 83.204 | 0.550 | 81.151 | 85.147 |
| | 50 | 92.061 | 0.342 | 90.890 | 93.379 |
| | 75 | 95.095 | 0.190 | 94.321 | 95.824 |

Table 4: NPCR statistics for JPEG.

| Quality | | $N_{0.05}^*$ | $U(O, \mathcal{Z}(O))$ | pass % |
|----------------|----|--------------|------------------------|--------------|
| <i>foreman</i> | 25 | 96.116 | 4.984 | 0 |
| | 50 | 96.205 | 5.107 | 0 |
| | 75 | 96.250 | 5.171 | 0 |
| <i>crew</i> | 25 | 95.985 | 4.813 | 0 |
| | 50 | 96.012 | 4.847 | 0 |
| | 75 | 96.051 | 4.897 | $< 10^{-10}$ |

Table 5: Critical values for the JPEG randomness test and the percentage of keys which passed the test.

randomaccess GOP structure should be disregarded since their key sensitivity is poor.

4.2 JPEG Analysis

ACSE is not able to provide content security for JPEG as shown in Figure 1. We analyze key and quality sensitivity using the same methodology as for H.265/HEVC. However, due to the lack of intra-frame dependencies in JPEG, zeroing ACCs of higher quality images will always result in a downscaled image of decent quality. Thus, we choose a quality level of 5 as a threshold indicator for sufficient encryption. At this quality level, DC coefficients as well as color information are severely degraded. As samples we use still frames of the *crew* and *foreman* sequences (frame 100) with quality levels 25, 50 and 75. As for H.265/HEVC, we test 100 different keys.

Key Sensitivity — The results are summarized in Tables 4 and 5. All samples fail the randomness test. However, the *crew* sequence with quality level 75 shows that for a busier picture and higher quality levels, the randomness test is passable. For quality level 95, $U(O, \mathcal{Z}(O)) = 4.939$ and consequently $N_{0.05}^* = 96.082$ which is passed by 99.998% of the keys with a NPCR of $\mu = 96.4$, $\sigma = 0.076$. For the *foreman* sequence, the randomness test is failed by all keys, even for a quality level of 95. In practice, one has to assume that the randomness test is failed unless shown otherwise.

Quality Sensitivity — The results of the quality sensitivity tests are summarized in Table 6. It can clearly be seen that the impact of ACSE for JPEG is much higher and all tested keys pass the sufficiency test. Furthermore, the standard deviation of the quality as well as the base average quality are relatively stable, meaning the JPEG ACSE is well suited for sufficient and transparent encryption. This is due to larger ACCs and consequently a higher local impact of the encryption. Adding to this is the missing intra-frame prediction of JPEG which means the

| | Quality | μ | σ | S^* | suff. % |
|---------|---------|--------|----------|--------|---------|
| foreman | 25 | 18.000 | 0.204 | 20.934 | 100.00 |
| | 50 | 17.862 | 0.215 | 20.934 | 100.00 |
| | 75 | 17.890 | 0.209 | 20.934 | 100.00 |
| creta | 25 | 20.851 | 0.167 | 24.406 | 100.00 |
| | 50 | 20.771 | 0.145 | 24.406 | 100.00 |
| | 75 | 20.692 | 0.140 | 24.406 | 100.00 |

Table 6: JPEG PSNR statistics, the threshold and percentage of keys which result in sufficient quality.

errors introduced by the encryption are not propagated further and the overall impact is much more stable.

5 Conclusion

We showed that encrypting only ACC signs does not allow for confidential encryption, neither for JPEG nor for H.265/HEVC, and thus for no other DCT-based standard in-between the two. Furthermore, we showed that sufficient and transparent encryption are possible for some, but not all encoding settings with a strong dependency on quality and the amount of motion. We also showed that the choice of the key used for encryption influences the security greatly. We recommend to use our analyses to evaluate the feasibility of ACC sign and similar encryption for a given use case.

References

- [1] T. E. Boult, "Pico: Privacy through invertible cryptographic obscuration," in *IEEE/NFS Workshop on Computer Vision for Interactive and Intelligent Environments*, Lexington, KY, USA, Nov. 2005, pp. 27–38 (cit. on p. 2).
- [2] P. Carrillo, H. Kalva, and S. Magliveras, "Compression Independent Reversible Encryption for Privacy in Video Surveillance," *EURASIP Journal on Information Security*, vol. 2009, pp. 1–13, Jan. 2009 (cit. on p. 2).
- [3] B. M. M. Macq and J.-J. Quisquater, "Cryptology for digital TV broadcasting," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 944–957, Jun. 1995 (cit. on p. 2).
- [4] H. Hofbauer, A. Uhl, and A. Unterweger, "Transparent encryption for hevc using bit-stream-based selective coefficient sign encryption," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy: IEEE, May 2014, pp. 1986–1990 (cit. on p. 2, 4).
- [5] C. Shi and B. Bhargava, "A fast MPEG video encryption algorithm," in *Proceedings of the Sixth ACM International Multimedia Conference*, Bristol, UK, Sep. 1998, pp. 81–88. [Online]. Available: [a%20href=%22http://www.acm.org/sigmm/MM98/electronic_proceedings/shi/index.html%22%3Eonline%20version%3C/a](http://www.acm.org/sigmm/MM98/electronic_proceedings/shi/index.html%22%3Eonline%20version%3C/a) (cit. on p. 2).
- [6] F. Dufaux and T. Ebrahimi, "Scrambling for Anonymous Visual Communications," in *Proceedings of SPIE, Applications of Digital Image Processing XXVIII*, vol. 5909, SPIE, 2005 (cit. on p. 2).
- [7] F. Dufaux and T. Ebrahimi, "H.264/AVC video scrambling for privacy protection," in *Proceedings of the IEEE International Conference on Image Processing, ICIP '08*, San Diego, CA, USA: IEEE, Oct. 2008, pp. 47–49 (cit. on p. 2).
- [8] L. Tong, F. Dai, Y. Zhang, and J. Li, "Prediction restricted H.264/AVC video scrambling for privacy protection," *Electronic Letters*, vol. 46, no. 1, pp. 47–49, Jan. 2010. doi: [10.1049/e1.2010.2068](https://doi.org/10.1049/e1.2010.2068) (cit. on p. 2).
- [9] Y. G. Won, T. M. Bae, and Y. M. Ro, "Scalable protection and access control in full scalable video coding," in *Proceedings on the 5th International Workshop on Digital Watermarking, IWDW '06*, ser. Lecture Notes in Computer Science, vol. 4283, Korea: Springer, Nov. 2006, pp. 407–421 (cit. on p. 2).
- [10] G. Van Wallendael, A. Boho, J. De Cock, A. Munteanu, and R. Van de Walle, "Encryption for high efficiency video coding with video adaptation capabilities," in *2013 IEEE International Conference on Consumer Electronics (ICCE)*, 2013, pp. 31–32 (cit. on p. 2).
- [11] G. Van Wallendael, J. De Cock, S. Van Leuven, A. Boho, P. Lambert, B. Preneel, and R. Van De Walle, "Format-compliant encryption techniques for high efficiency video coding," in *2013 20th IEEE International Conference on Image Processing (ICIP)*, Sep. 2013, pp. 4583–4587 (cit. on p. 2).
- [12] H. Hofbauer and A. Uhl, "Selective encryption of the MC EZBC bitstream for DRM scenarios," in *Proceedings of the 11th ACM Workshop on Multimedia and Security*, Princeton, New Jersey, USA: ACM, Sep. 2009, pp. 161–170 (cit. on p. 2).
- [13] H. Hellwagner, H. Hofbauer, R. Kuschnig, T. Stütz, and A. Uhl, "Secure transport and adaptation of MC-EZBC video utilizing H.264-based transport protocols," *Elsevier Journal on Signal Processing: Image Communication*, vol. 27, no. 2, pp. 192–207, 2011, issn: 0923-5965. doi: [10.1016/j.image.2011.11.002](https://doi.org/10.1016/j.image.2011.11.002). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0923596511001330> (cit. on p. 2).
- [14] Y. Wu, J. P. Noonan, and S. Aghaian, "NPCR and UACI randomness tests for image encryption," *Journal of Selected Areas in Telecommunications*, pp. 31–38, Apr. 2011. [Online]. Available: <http://www.cyberjournals.com/Papers/Apr2011/05.pdf> (cit. on p. 3).
- [15] ITU-T T.81, *Digital compression and coding of continuous-tone still images — requirements and guidelines*, Also published as ISO/IEC IS 10918-1., Sep. 1992 (cit. on p. 3).
- [16] ITU-T H.265, *High efficiency video coding*, <http://www.itu.int/rec/T-REC-H.265-201304-I>, Apr. 2013 (cit. on p. 3).