Selective Encryption for Hierarchical MPEG

Heinz Hofbauer, Thomas Stütz*, and Andreas Uhl

University of Salzburg,
Department of Computer Sciences
{hhofbaue, tstuetz, uhl}@cosy.sbg.ac.at

Abstract. Selective encryption of visual data and especially MPEG has attracted a considerable number of researchers in recent years. Scalable visual formats are offering additional functionality, which is of great benefit for streaming and networking applications. The MPEG-2 and MPEG-4 standards provide a scalability profile in which a resolution scalable mode is specified. In this paper we evaluate a selective encryption approach on the basis of our hierarchical MPEG video codec.

1 Introduction

Encryption schemes for visual data need to be specifically designed to protect the content while preserving properties of its representation in the encrypted domain. Furthermore the real-time encryption of a video stream with state-of-the-art ciphers still requires heavy computation, especially when considering target hardware platforms like set-top boxes for digital TV broadcasts. Numerous attempts have been made to secure MPEG streams, among them [1–4]. Selective encryption has been accomplished in various ways, encryption of I-frames, motion vector data, coefficient permutation, ... Several approaches do not strive for maximum security, but trade off security for computational complexity. For a detailed discussion please refer to [5]. Also the JPEG standard and its scalable modes of operation were target of research concerning selective encryption [6, 7]. The rising importance of scalability preserving encryption of scalable video streams has been discussed in [8–10]. Selective encryption of scalable video streams can greatly reduce the complexity of video distribution in different qualities/resolutions. If the scalability is preserved in the encrypted domain, no decryption key, no decryption, no transcoding and no reencryption is necessary for accessing the lower resolution versions of a stream. The high quality layers can simply be dropped e.g., by a simple network set-top box. The computational effort is to parse the code stream for marker sequences to identify the relevant parts, which is negligibly small.

The paper presents a selective encryption approach for a hierarchical MPEG coder and evaluates its performance. An advantage of using scalable video codecs is that the overhead to identify the relevant parts for encryption can be kept very low and that rate adaption can easily be conducted in the encrypted domain.

* The support of the Austrian Grid project is gratefully acknowledged.

In section 2 we give an overview of the hierarchical MPEG coder. The selective encryption approach is presented in section 3. Empirical results of this selective encryption approach are discussed in section 4 and section 5 concludes the paper.

2 Hierarchical MPEG

HMPEG (Hierarchical MPEG) is closely related to the MPEG resolution scalable mode as defined for MPEG-2 and MPEG-4 Part 2. Since there is no freely available implementation capable of compressing in this mode (the scalable modes are altogether poorly supported) we had to implement it from the scratch. The compression performance of the HJPEG coder is very close to baseline JPEG [6]. The MPEG-2 code stream syntax is rather complex and of no special interest for our investigations, thus our implementation does not produce a standard-compliant MPEG-2 stream but has all its properties. Hence all our results are also applicable to resolution scalable MPEG-2. The MPEG standards basically apply motion compensation to exploit temporal redundancy and compress the resulting frames in a way very similar to JPEG. Our implementation employs hierarchical JPEG as defined in [11] for frame compression.

2.1 Hierarchical JPEG

HJPEG (Hierarchical JPEG) is a resolution scalable compression method, which is part of the JPEG standard [11]. A number of layers is chosen and for each layer the image is downsampled by a factor of two in each dimension with upsampling and downsampling filters as proposed in [11]. The reconstruction of a certain resolution is used as a prediction for the next higher resolution. The resulting series of difference frames and the lowest resolution subsample are JPEG encoded. The HJPEG compression is conducted with a custom implementation based on the Independent JPEG Group's library libjpeg.

2.2 MPEG Video Compression

For motion compensation a frame is split up into macroblocks (16x16 pixel in MPEG-2 and our implementation). For each of these macroblocks the best matching block in another frame is located (in our implementation with full pixel accuracy) and the difference calculated. Additionally there is the possibility that no good enough macroblock can be found, resulting in a so called I-macroblock which contains original image information. Motion compensation is accompanied by a still image compression system that applies a DCT (discrete cosine transform) and a Huffman based entropy coder. MPEG-2 uses three different frame types:

- I-Frame (Intra Frame): contains solely original image information.
- P-Frame (Predicted Frame): contains the difference between the previous and the actual frame.

– B-Frame (Bidirectionally Predicted Frame): uses the previous and next I- or P-frame for the computation of the prediction.

The repeated structure of these frames is called group of pictures (GOP) and has to start with an I-frame, e.g., I B B P B B P.

Putting all together our HMPEG implementation performs motion compensation on the basis of 16x16 pixel macroblocks with a one pixel accuracy. HJPEG is used to compress the frames in a resolution scalable fashion delivering a resolution scalable video stream.

3 Selective Encryption

One goal of selective encryption is the preservation of the scalability in the encrypted domain. Thus no key, no decryption, no transcoding and reencryption is necessary to access lower resolution versions of the video stream, as it would be the case for regular encryption and non-scalable video data. Another possible goal is the reduction of the encryption complexity by reducing the amount of data to be encrypted. Application scenarios for selective encryption can be divided into two groups.

Confidential encryption has the same goal as regular encryption, the secure scrambling of all image information.

Transparent encryption is here used as an umbrella term for all application scenarios where confidential encryption is not demanded. These application scenarios may impose two requirements:

– Security requirement: a certain portion of the visual information has to be securely removed.
– Quality requirement: a certain image quality may have to be preserved.

A content provider might want to reveal a low quality version in order to attract costumers and lure them into buying the high quality version. Another case is e.g., the transmission of a soccer game. The information that a soccer game is broadcasted is not subject to secrecy, but nevertheless the broadcast should only be consumable by paying customers, that can decipher the encrypted broadcast. Here the security requirement is to achieve sufficiently bad quality.

3.1 Approach

The targets of our encryption approach are the compressed coefficient data of the frames and the motion vectors. Therefore we have to identify these parts. In our implementation we had to deal with the JPEG code stream syntax but essentially the same is possible with MPEG-2 code stream syntax. For the JPEG code stream syntax we have to parse for markers indicating a scan, the JPEG code stream unit containing only compressed coefficient data. After encryption with a state-of-the-art cryptographic cipher (AES [12]) we apply byte stuffing to avoid maker sequences and to preserve the scalability. The code stream syntax is

at least partially preserved except for valid Huffman codes. In the case of MPEG the parsing of the relevant parts is different. In MPEG-2 these marker sequences are called start codes and consist of a 3 byte prefix (0x00 00 01) followed by a 1 byte identifier. The relevant parts (compressed coefficient data and motion vectors) have to be identified (slices in MPEG) and the generation of start codes prevented.

3.2 Attacks

The proposed encryption approach results in a pseudo code-stream-compliant file which can be at least partly recovered by a code-stream-compliant decoder. However, this reconstruction will have a significantly lower quality than a possible attacker can achieve. The reason is that our method introduces heavy random

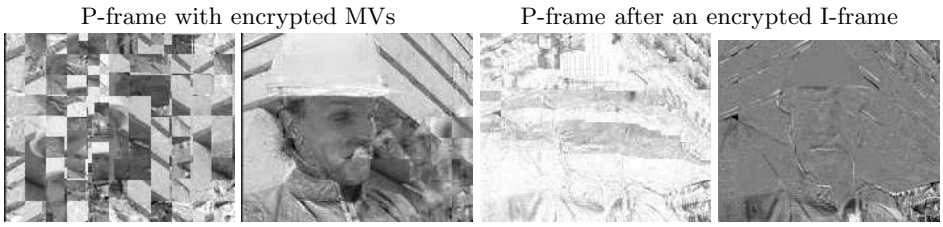

Fig. 1. Replacement attacks versus the reconstruction of the partially encrypted Foreman video.

distortion which also greatly reduces the image quality of unencrypted parts of the video stream. An attacker can replace encrypted parts with data that does not introduce random noise. This method is called replacement attack. In our case encrypted I-frames are replaced by a uniform gray image. The encrypted difference frames are replaced by zero valued frames.

If the motion vectors are encrypted the replacement attack is to either set them to $(0, 0)$ or to decode each difference frame to obtain the high frequency changes between frames. Figure 1 illustrates the replacement attack for various coding and encryption settings.

4 Results

HMPEG with two HJPEG layers was applied and all results are obtained by conducting a replacement attack. Two layers were used because the resolution of the Foreman test sequence has been too low (176x144) to justify more layers. In this paper only the PSNR plots are given, but also the LSS/ESS values [13] were evaluated. The LSS/ESS plots basically show the same behavior as the PSNR plots and therefore only the more familiar PSNR plots are given.

4.1 Confidential Encryption

All HJPEG layers contain at least high frequency data that reveals information about the image content. Figure 2 reveals that this also holds for predicted frames (only a minor adjustment of the color levels was performed). Therefore - and on the basis of many other evaluations - one can state that nearly all of the HMPEG stream has to be encrypted to achieve perfect secrecy. Even the motion vectors alone reveal enough image information to roughly guess the content.

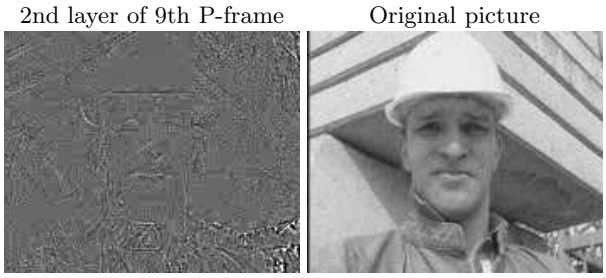

Fig. 2. Image information in a HJPEG layer of a predicted frame

4.2 Transparent Encryption

The requirements of transparent encryption may vary from application to application. The following results will hopefully give the reader some insight into what can be achieved with selective encryption of HMPEG. An application that needs to preserve a certain image quality only keeps the resolutions unencrypted that satisfy the quality requirement. In the following we tried to analyze how the encryption effort can be minimized while severely distorting the video quality. We evaluated the image quality for various encryption and coding settings, including different GOP structures.

Figure 3 shows the PSNR plot of the first 125 frames (GOP: one I-frame and the rest P-frames) of the Foreman sequence with various parameter and encryption settings. Only the I-frame has been encrypted (both layers). Encrypted motion vectors are indicated by the label `mv` and the usage of I-macroblocks is indicated by the label `imb`. To take into account that the addition of difference frames without a reference I-frame may also introduce additional distortion, we also plotted the PSNR of the layer 1. This layer 1 PSNR plot (`layer1direct`) extracts the highest resolution HJPEG layer of each frame and uses it directly. Since the highest resolution HJPEG layer produces the biggest amount of HMPEG data, encrypting it leads to the encryption of most of the HMPEG stream. In this sense the `layer1direct` PSNR graph shows a lower bound for the video quality while only encrypting a rather small portion of the overall HMPEG stream data.

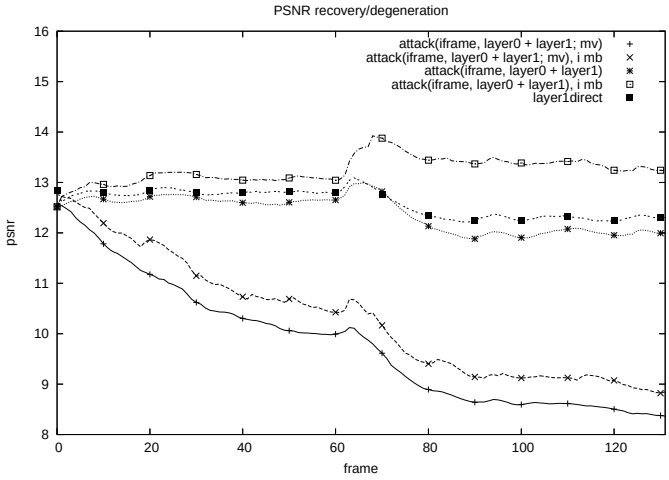

Fig. 3. PSNR of the HMPEG compressed Foreman sequence for a GOP with only P-frames, I-frame encrypted.

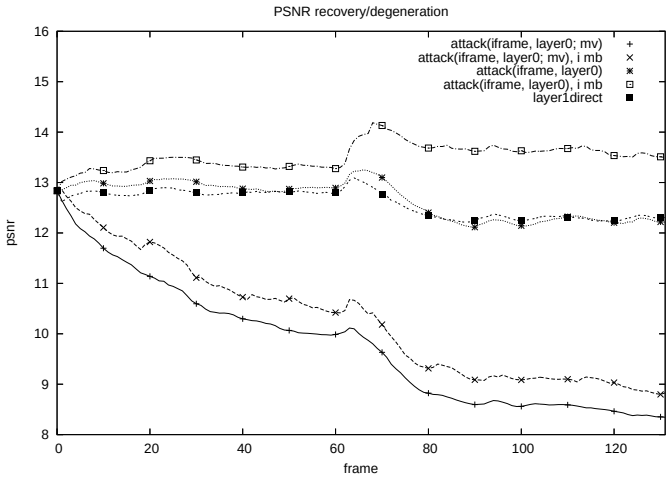

Fig. 4. PSNR of the HMPEG compressed Foreman sequence for a GOP with only P-frames, base layer of I-frame encrypted.

The original image, an estimation of the image via replacement attack of the I-frame (no I-macroblocks, unencrypted motion vectors) and the direct use of layer 1 is illustrated in Fig. 5 for the 100th P-frame of the Foreman sequence. Furthermore, should the PSNR of a sequence generated by a replacement attack be lower than the `layer1direct` sequence the partial encryption is sufficient not

Original frame Replacement attack Layer 1 direct

Fig. 5. Reconstructions based on two attacks compared to the original image for the Foreman sequence for the 100th P-frame (GOP with only I-frames)

only to remove all information of the I-frame but also to prevent any noticeable regenerative effect induced by the P-frames. Figure 3 states that the encryption of the I-frame is sufficient for destroying the visual quality of the whole sequence. Furthermore, when the motion vectors are encrypted the quality decreases even more, due to the addition of the difference frames to wrong spatial locations of the image. Figure 6 illustrates the restricted plausibility of the PSNR. The 50th P-frame of the `layer1direct` sequence, the sequence with the I-frame encrypted and the sequence with I-frame and motion vectors encrypted. Encrypting and attacking the motion vectors really degrades the quality as the PSNR plot would suggest.

Nevertheless if the motion vectors are left unencrypted, the quality of the image is better than its PSNR value indicates. While the PSNR of the attacked frame is lower than the PSNR of the layer 1 of the frame, certain details are better visible in the attacked frame and there is even some regeneration regarding the texture.

Layer 1 direct Encrypted I-frame Encrypted I-frame & MVs

Fig. 6. The 50th P-frame of the Foreman sequence for different encryptions and attacks, no I-macroblocks are used (GOP with only P-frames).

Regardless of these problems the PSNR plot of the Foreman sequence is a typical one. Unsurprisingly the use of I-macroblocks raises the quality of the frames, though without changing the basic nature of the results. The encryption of the motion vectors helps to prevent the sequence from regenerating and `layer1direct` yields to a better result than the replacement attack. The sequence with only the I-frame encrypted is better than the layer 1 used directly, although the plot suggests otherwise. Interestingly the encryption of only the base layer of an I-frame yields to very similar results. (see Fig. 4). This means only a fraction of the I-frames data needs to be encrypted to degenerate the quality in a way quite similar to encrypting the whole I-frame. The amount of movement in the sequence influences the gain of image quality when using I-macroblocks. In sequences where there is little movement, I-macroblocks will seldom come to bear since the difference is mostly caught by the motion vectors. Such a case is illustrated in Fig. 7 for the Akiyo sequence. If the GOP is

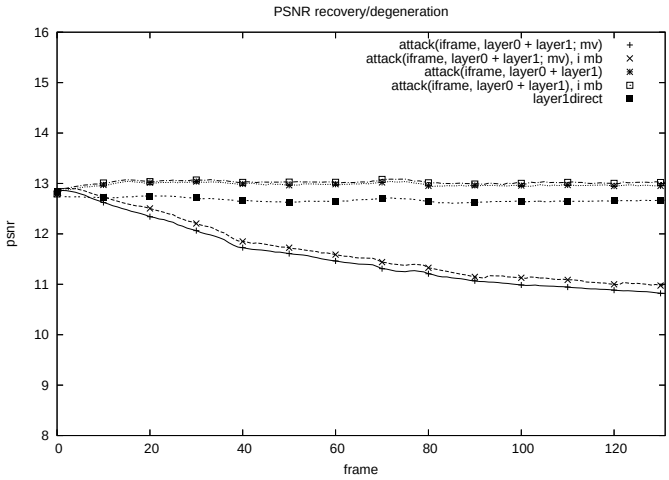

Fig. 7. PSNR of the HMPEG compressed Akiyo sequence with a GOP of only P-frames, I-frame encrypted.

changed to IBBPBBP... the overall behavior stays the same but the quality becomes slightly better. This is due to the higher distance between P-frames, which results in difference frames containing more information. For the same reason the influence of I-macroblocks with this GOP is higher, since the greater difference between the frames leads to a higher number of I-macroblocks. Figure 8 shows a PSNR plot for the Foreman sequence using a GOP of IBBPBBP.... Tough PSNR would suggest a noticeable increase in the visual quality this higher PSNR only reflects the higher number of I-macroblocks. For an IBBPBBP... GOP using I-macroblocks the 50th image of an attacked sequence is depicted in Fig. 9.

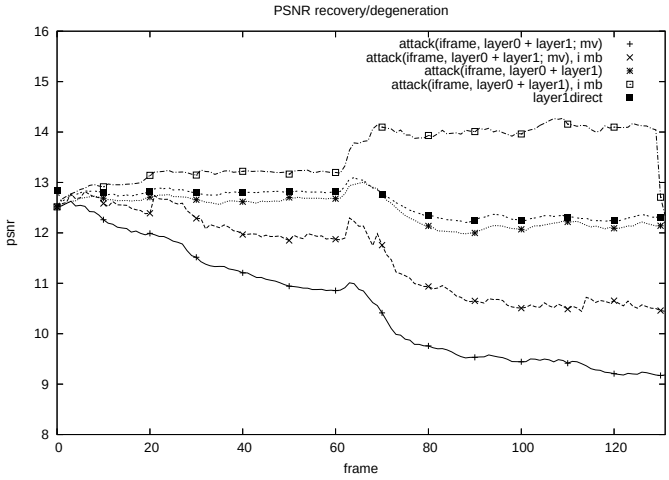

Fig. 8. PSNR of the HMPEG compressed Foreman sequence with a GOP of IBBPBBP. . ., I-frame encrypted .

Layer 1 direct | Encrypted I-frame | Encrypted I-frame & MVs

Fig. 9. The 50th P-frame of the Foreman sequence for different encryptions and attacks, I-macroblocks are used (GOP with P- and B-frames) .

The visual quality is still poor and has not changed much, except for the visible I-macroblocks, regarding a GOP with one I-frame and the rest P-frames (c.f. Fig. 6).

Again very similar results are obtained by encrypting only the base layer of the I-frame.

5 Conclusion

In this paper we presented a selective encryption approach for HMPEG that is capable of preserving the scalability in the encrypted domain. Furthermore we evaluated its suitability for confidential and transparent encryption. For confidential encryption most of the video stream has to be encrypted including

motion vectors. Nevertheless the scalability is fully preserved even on a JPEG scan basis for our implementation or on a slice basis for MPEG. Furthermore we could show that it is possible to severely degrade the video quality by only encrypting very little of the overall video stream data (about 0.3% for 125 P- or B-frames in a GOP). It is sufficient to encrypt the base layer of all I-frames in order to severely distort and reduce the video quality.

References

1. Kunkelmann, T.: Applying encryption to video communication. In: Proceedings of the Multimedia and Security Workshop at ACM Multimedia '98, Bristol, England (1998) 41–47
2. Dittmann, J., Steinmetz, R.: A technical approach to the transparent encryption of MPEG-2 video. In Katsikas, S.K., ed.: Communications and Multimedia Security, IFIP TC6/TC11 Third Joint Working Conference, CMS '97, Athens, Greece, Chapman and Hall (1997) 215–226
3. Bhargava, B., Shi, C., Wang, Y.: MPEG video encryption algorithms. Multimedia Tools and Applications **24**(1) (2004) 57–79
4. Qiao, L., Nahrstedt, K.: Comparison of MPEG encryption algorithms. International Journal on Computers and Graphics (Special Issue on Data Security in Image Communication and Networks) **22**(3) (1998) 437–444
5. Uhl, A., Pommer, A.: Image and Video Encryption. From Digital Rights Management to Secured Personal Communication. Volume 15 of Advances in Information Security. Springer-Verlag (2005)
6. Stütz, T., Uhl, A.: Image confidentiality using Progressive JPEG. In: Proceedings of the International Conference on Information, Communications & Signal Processing, ICICS '05, Bangkok, Thailand (2005)
7. Fisch, M.M., Stögner, H., Uhl, A.: Layered encryption techniques for DCT-coded visual data. In: Proceedings (CD-ROM) of the European Signal Processing Conference, EUSIPCO '04, Vienna, Austria (2004) paper cr1361.
8. Wee, S., Apostolopoulos, J.: Secure scalable video streaming for wireless networks. In: Proceedings of the 2001 International Conference on Acoustics, Speech and Signal Processing (ICASSP 2001), Salt Lake City, Utah, USA (2001) invited paper.
9. Wee, S., Apostolopoulos, J.: Secure scalable streaming enabling transcoding without decryption. In: Proceedings of the IEEE International Conference on Image Processing (ICIP'01), Thessaloniki, Greece (2001)
10. Wee, S., Apostolopoulos, J.: Secure scalable streaming and secure transcoding with JPEG2000. In: Proceedings of the IEEE International Conference on Image Processing (ICIP'03). Volume I., Barcelona, Spain (2003) 547–551
11. Pennebaker, W., Mitchell, J.: JPEG – Still image compression standard. Van Nostrand Reinhold, New York (1993)
12. National Institute of Standards and Technology: FIPS-197 - advanced encryption standard (AES) (2001)
13. Mao, Y., Wu, M.: Security evaluation for communication-friendly encryption of multimedia. In: Proceedings of the IEEE International Conference on Image Processing (ICIP'04), Singapore, IEEE Signal Processing Society (2004)