No Tradeoff between Confidentiality and Performance: An Analysis on H.264/SVC Partial Encryption

Zhuo Wei, Xuhua Ding, Robert Huijie Deng, Yongdong Wu

No Tradeoff between Confidentiality and Performance: an Analysis on H.264/SVC Partial Encryption

Zhuo Wei[1], Xuhua Ding[1], Robert Huijie Deng[1], Yongdong Wu[2]

[1] School of Information Systems, Singapore Management University, 178902
phdzwei@gmail.com, {xhding, robertdeng}@smu.edu.sg
[2] Institute for Infocomm Research, 1 Fusionopolis Way Singapore 138632
wydong@i2r.a-star.edu.sg

Abstract. Partial encryption is often used as a tradeoff between security and performance to protect scalable video data. In this paper, we argue that although partial encryption is strong enough for access control, it is not adequate for content confidentiality protection. We conduct experiments to show that partially encrypted H.264/SVC (scalable video coding) streams leak significant content information from the enhancement layers in all three scalability dimensions. Our analysis concludes that such leakage is caused by the underlying coding techniques used in H.264/SVC, and all layers should be encrypted to protect confidential video streams.

Keywords: Scalable Video Coding, H.264/SVC, Partial Encryption.

1 Introduction

Scalable video streaming techniques, such as MPEG-4 FGS (fine grain scalability) [1] and H.264/SVC (scalable video coding) [2, 3], are widely used in real time content distribution due to their adaptability to a variety of heterogenous network and platform settings. Accompanying the growth of such techniques is the conflicting requirements between the protection of content confidentiality and the demand for lightweight computation on the content sender and receivers. *Partial encryption* or *selective encryption* is one of the widely adopted approaches to strike a balance between security and performance. Examples of partial encryption techniques include [4–14]. In contrast to *full encryption* algorithms whereby all content data are encrypted, *partial encryption* algorithms only encrypt those data which are considered important, e.g., the SVC base layers or Intra-coded blocks, while ignore other data. By reducing the amount of encryption operations, partial encryption algorithms aim to reduce the encryption overhead without undermining security. Several works [15–18] have discussed the security of partial encryption for MPEG-4 with the focus on the temporal layers. A comprehensive review on H.264/AVC (advanced video coding) encryption is given in [19]. Generally, SVC encryption finds two kinds of applications: transparent

encryption (or access control) and confidentiality. The former subjectively leaves low quality video data for users' previewing while the latter prevents the exposure of potentially sensitive material (e.g., identification of people, objects, data, and so on) on the entire video data. In this paper, we focus on the latter type of applications of partial encryption. In this kind of partial encryption, the argument is that those data such as enhancement layers are left in plaintext since they do not leak sensitive information as long as the adversaries cannot decrypt the base layers. We systematically investigate the security of partial encryption for H.264/SVC from all three scalability dimensions, i.e. the spatial, quality and temporal scalability. Our experimental results show that partial encryption fails to strike the desired balance because its does not offer satisfactory security strength for confidentiality protection. To gain more insights, we further investigate the relationship between confidentiality and scalability of H.264/SVC in the light of scalable coding techniques, and conclude that all layers have to be encrypted for confidential video streams.

The rest of our paper is organized as follows. Section 2 reviews partial encryption of SVC, this is then followed by our partial encryption experiments given in Section 3. Section 4 presents theoretical analysis and objective evaluation of leakage for SVC partial encryption. Section 5 introduces related work. Finally, we conclude our paper in Section 6.

2 Review on Partial Encryption

Partial encryption is to preserve multimedia property (e.g., format-compliance, scalability) by treating different data in a multimedia stream differently according to their importance. The basic idea is that those critical data are encrypted rigorously whereas those non-critical data are weakly protected or even not protected with the hope that the overall security strength is still maintained. The implementation of this idea varies with the scalable media type and the dimension of scalability as described in this section.

Scalable video coding includes wavelet-based SVC, MPEG-4 FGS, and H.264 SVC. Based on the granularity of scalability leveraged by the partial encryption schemes, we classify them into spatial/quality and temporal levels.

On Spacial and Quality Scalability Encryption algorithms in this category treat the base layer and the enhancement layers (typically in spatial/quality scalability) differently. For wavelet-based SVC, a subband-adaptive approach to scramble surveillance video content (scalable video coding with JPEG XR) is proposed in [4, 5], which scrambles DC and LP (low pass) subbands, but only inverts the signs of coefficients for HP (high pass) subbands and leaves Flexbits subbands in plaintext.

Unlike wavelet-based SVC, MPEG-4 FGS and H.264/SVC bitstreams are composed of a base layer and one or multiple scalable enhancement layers. Partial encryption algorithms for MPEG-4 FGS and H.264/SVC typically apply a strong cipher for the base layer, and use selective encryption or even no encryption for

the enhancement layers. For instance, in [6], the based layer is encrypted by the Chain and Sum cipher and the sign bits of DCT (discrete cosine transform) coefficients in enhancement layer are masked with a random sequence generated by RC4 [20]. The schemes in [21–23] encrypt an H.264/SVC base layer's intra prediction modes, the motion vector difference values and the sign bits of the texture data, whereas only the texture sign bits and the MVD (motion vector difference) sign bits in the spacial and quality enhancement layers are encrypted.

Quality scalability can also be achieved using DCT coefficients whereby low frequency coefficients represent the base layer and the middle or high frequency coefficients represent the enhancement layers. The idea of partial encryption is realized by encrypting DC or low frequency AC coefficients while the high frequency AC coefficient encryption being dismissed. For example, as proposed in [10], the first five coefficients and the subsequent fifteen coefficients are encrypted as the base layer and the middle layer respectively, while the remaining coefficients are in plaintext as high layer.

On Temporal Scalability A compressed video sequence is composed of I, P, and B frames, where the latter two are temporal enhancement layers in scalable video coding. Partial encryption algorithms at the temporal scalability are based on the observation that P frames and B frames are not meaningful when rendered without the corresponding I-frame. Typically, this type of algorithms provides a strong security for I frames while ignoring P and B frames.

For example, Hong et al. in [24] propose an encryption scheme for temporal scalable video coding whereby the motion vectors and residual coefficients of P or B frames are in plaintext. Meanwhile, Li et al. in [11] propose an encryption scheme for H.264/SVC at the NAL (network abstraction layer) level. For all NAL units, Instantaneous Decoding Refresh (IDR) Picture, Sequence Parameter Set (SPS), and Picture Parameter Set (PPS) are encrypted with a stream cipher. However, it has no protection over other temporal enhancement NALs.

3 Experiments of Partial Encryption

To systematically understand the security implication of partial encryption, we design a series of experiments for H.264/SVC, and evaluate the partial encrypted video streams.

3.1 Scalable Video Experiments

In our scalable video experiments, we choose ten standard benchmark video sequences[3] in order to cover different combinations of video characteristics including motion (fast/slow, pan/zoom/rotation), color (bright/dull), contrast (high/low), and object type (vehicle, buildings, people). *Bus* and *Foreman* video sequences

[3] Available at http://media.xiph.org/video/derf/

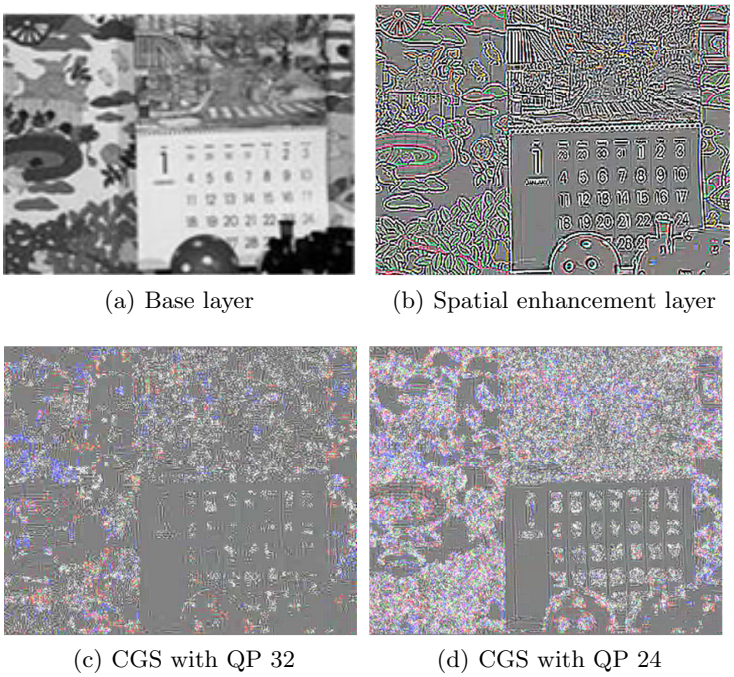
are of no camera motion while *Football* and *Soccer* demonstrate camera panning and zooming with object motion and background texture. *Bridge-far* and *Bridge-close* show images with smooth motion. *Highway* is a sequence of fast motion while *Silent* is a static sequence except of a person's right hand. *Mobile* and *Hall* sequences display a still complex background with foreground motion. All these sequences are encoded with the temporal, spatial and quality enhancement layers. Each GOP (group of picture) includes 16 frames and the I-frame Interval is set as 32.

The main rationale of our experiments is that given an SVC video sequence in plaintext, we strip off the base layers to simulate the effect that an adversary acquires no semantic information from a properly encrypted based layer. Then, we decode the remaining SVC enhancement layers using the default prediction mode, and check whether they leak semantic information about the video. We also apply certain weak encryption (e.g., sign encryption) to the enhancement layers and check the leakage from the ciphertext. Our experiments are implemented with JSVM 9.19 [25].

Spatial Scalability The spatial enhancement layer utilizes inter-layer prediction mechanisms [2] in order to increase compression efficiency. It only transmits the residual signals. In the spatial scalability experiments, we set the frames of the base layer as blank when decoding the enhancement layers. For all ten sequences in testing, I-frames of the enhancement layers are decoded, and they all reveal sufficient texture information of the objects in the sequences. For example, Figure 1(b) illustrates content leakage of enhancement layer for the *Mobile* sequence. The ten experimental results also indicate some texture are easier to recognize, such as face, non-overlap objects, and the leakage becomes more evident when the video stream (only containing enhancement layers) is played.

Quality Scalability When using inter-layer prediction for CGS (coarse grain scalability) of H.264/SVC, a refinement of texture information is typically achieved by requantizing the residual texture signal of the enhancement layer with a smaller quantization step size relative to that used in the reference layer. In our CGS scalability experiments, the images of the reference layer are set as blank, meanwhile, motion vectors are all set as zero. The experiment sets QP (quantization parameter) of the base layer as 34 and sets QP of the enhancement layers from 24 to 32. Richer and non-overlap texture of images can easily produce leakage in the enhancement layer. QP difference also affects the amount of leakage. Sensitive contents of all ten sequences can be detected if the decoded images of the enhancement layer are continuously played. Figure 1(c) and 1(d) show data leakage of *Mobile* for QP 32 and QP 24.

Partial encryption on MGS (medium grain scalability) enhancement layers is not secure either, though the layering techniques are different with CGS. MGS layers are generally parts of CGS as MGS and CGS take the same QP. Each MGS layer is composed of part frequency coefficients of 4×4 DCT to supplement quality enhancement for base layer. We set the transform coefficients with different MGS layers based on the zigzag order (important and unimportant). In



(a) Base layer (b) Spatial enhancement layer

(c) CGS with QP 32 (d) CGS with QP 24

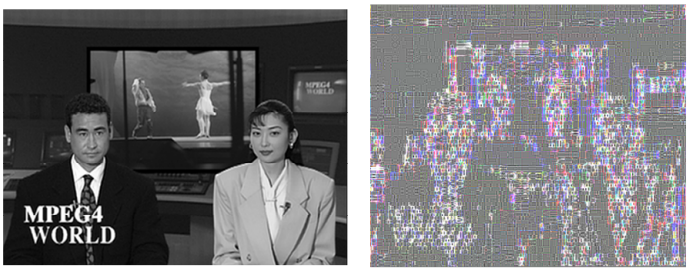
Fig. 1. Experiment as spatial and CGS scalability layers for *Mobile* sequence.

our MGS experiments, there are three MGS layers: (1) MGS0: first three coefficients, set other coefficients with zero; (2) MGS1: the 3rd to the 5th coefficients, set other coefficients with zero; (3)MGS2: the 6th to the 15th coefficients, set other coefficients with zero.

Similar to CGS, the QP difference between the enhancement layer and the reference layer affects the amount of disclosure. Moreover, the content of each MGS layer is related to non-zero coefficients which depend on texture feature. For example, Figure 2 illustrates the first decoded images corresponding to different MGS layers for *News* sequence. For all MGS layers, the profile of two speakers are apparent and the dancers on the TV can be easily viewed when being continuously played.

Temporal Scalability The temporal enhancement layer depends on Inter prediction technique, which uses a range of block sizes from 16×16 to 4×4 to predict pixels in the current frame from similar regions in previously frames. These previously coded frames may occur before or after the current frame in display.

We encode the ten video sequences with four temporal scalability layers. We set the images between the temporal layer 0 to the temporal layer 2 as blank and define motion vectors as zero, then only decode the temporal layer 3. Experi-



(a) Base layer (b) MGS0

(c) MGS1 (d) MGS2

Fig. 2. Experiment as MGS scalability layers for *News* sequence. (a) is original image of base layer; (b) to (d) are MGS layers corresponding to images which only contain parts of coefficients of enhancement layers.

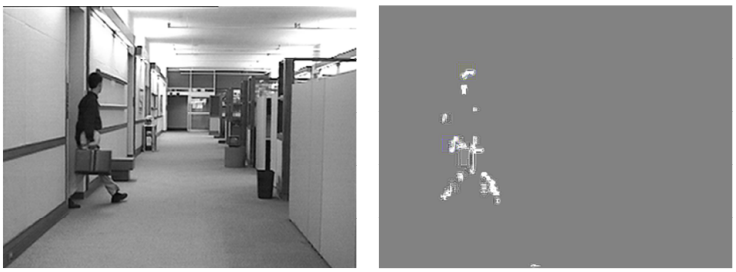
mental results indicate that the temporal enhancement layer of motion sequences can cause significant information leakage. However, the temporal enhancement layers of static sequences, such as *Bridge-close*, *Bridge-far*, generally have less residuals and cause less leakage. Figure 3(a) illustrates the 26th frame of *Hall* sequence in which the person's profile can clearly detected.

4 Theoretical Analysis and Objective Assessment

Our discussion below focuses on H.264/SVC. Nonetheless, other scalable video coding standards, such as MPEG-4 FGS, share the same prediction coding techniques with H.264/SVC, e.g., in terms of prediction, DCT, quantization, and entropy coding. Therefore, our results are applicable to all scalable video encoding standards.

4.1 Leakage Detection

The leakage of a video stream can be identified on spatial texture contour or/and temporal motion objects residual, which carry the semantic information about the objects in the stream. Although the leakage can be visible to human eyes



(a) Base layer (b) Temporal enhancement layer

Fig. 3. Experiment as temporal scalability layers for *Hall* sequence.

and therefore are identified manually, they can be also measured by using edge detection and matching techniques. Leakage on resolution and quality scalability is associated with texture enhancement. Images rendered by the enhancement layers alone share the same contour with the original one displayed from the entire video stream. We use SOBEL edge detection [26] on both the original and the enhancement layer image to obtain the contour. Then, we compare two sets of edges and derive an edge similarity score (ESS) [27] which measures the degree of resemblance of the edge and contour information between images.

Temporal scalability leakage is related to motion objects. When temporal frames are viewed as a video sequence, the outlines and trajectories of moving objects are readily visible. Similar to detect quality leakage, we utilize SOBEL to detect the edges and compute ESS scores for the similarity of moving objects or Intra-coded block areas between original and temporal enhancement images.

In practice, 0.5 is chosen as the safety ESS threshold as suggested in [27] for using encryption for access control purpose. For those sensitive applications demanding confidentiality, it is desirable that the ESS score should be close to zero, indicating that an encrypted frame does not leakage information about the plaintext.

4.2 Scalable Video Coding

The leakage shown in Section 3 is not by coincidence. In fact, it is the coding techniques used in scalable video that determines the content leakage from the enhancement layers.

Spatial Scalability Figure 4(a) illustrates the coding flow of spatial scalability. In spatial scalability encoding, an encoder first reconstructs the frame ($frame'_{bl}$) of a lower layer and upsamples it to produce a reference frame ($frame_{ups}$) that has the same effective resolution as the enhancement layer. $frame_{ups}$ is then used to generate the residual signal frame $frame^{res}_{el}$ for spatial enhancement layer before the compression. Therefore, the amount of content leakage from a

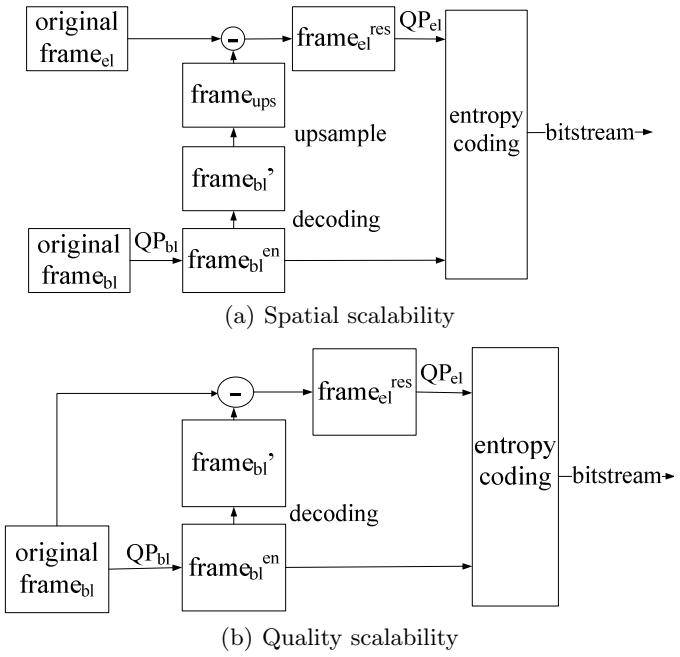


Fig. 4. The coding flow of spatial and quality scalability under inter-layer prediction technique.

spatial enhancement layer is determined by the amount of residue signals, which is relevant to the smoothness of the macroblocks in $frame_{bl}^{'}$ and the quantization step size.

The upsampling technique generally uses an one-dimensional 4-tap FIR (finite impulse response) filter for luminance components and a bilinear filter for chroma components, which involves visually disturbing signal components. If the macroblocks of $frame_{bl}^{'}$ are smooth, the upsampling technique can provide a well interpolation prediction for the corresponding macroblock of $frame_{el}$ due to similar information. The difference of the corresponding macroblocks between the $frame_{el}$ and the $frame_{ups}$ are small and less residual information are needed. If the macroblocks of $frame_{bl}^{'}$ contain rich texture and/or edge features, the subblocks of these macroblocks are independ with each other. For these kinds of macroblocks, upsampling introduces noise signals in the $frame_{ups}$ because irrelevance neighbors information are used for interpolation. Therefore, the difference of corresponding macroblocks between the $frame_{el}$ and the $frame_{ups}$ are large, which demands more residue information.

The quantization step sizes also affect the amount of in the enhancement layer. For a small QP_{bl}, $frame_{bl}^{'}$ has less distortion from quantization compression so that the difference between $frame_{bl}^{'}$ and $frame_{bl}$ is also small. Consequently, the $frame_{ups}$ will be more similar to the $frame_{el}$ so that the smooth

areas of $frame_{el}^{res}$ contains little residual information. In addition, the amount of residual information is also related to the gap between QP_{el} and QP_{bl} result in more residual data in $frame_{el}^{res}$. The larger the gap, the more amount residue information is in the enhancement layer.

Based on our spatial experiments, we also calculate the number of non-zero coefficients of base and enhancement layers. Entropy statistics indicates that the number of non-zero coefficients of spatial enhancement layer is about eight times of that of the base layer. In addition, Table 1 illustrates the ESS evaluation scores for ten video sequences. Six of them are even higher than the threshold used in access control. It is evident that the contour of the $frame_{el}^{'}$ has a strong similarity with the $frame_{el}$. In other words, those enhancement layers disclose the visual texture information about the objects in the video stream.

Table 1. ESS score of the $frame_{el}^{'}$ against the $frame_{el}$ under edge detection

Sequences	*Foreman*	*Hall*	*Bridge* -close	*Highway*	*Mobile*	*Silent*	*Soccer*	*Bridge* -far	*Football*	*Bus*
ESS	0.51	0.46	0.54	0.39	0.60	0.40	0.57	0.55	0.48	0.52

Quality Scalability Figure 4(b) illustrates the coding flow of quality scalability. We consider CGS and MGS separately in our analysis.

Coarse Grain Scalability CGS utilizes Inter-layer prediction without upsampling, because the layers of SVC are of the same resolution. The reconstructed frame ($frame_{bl}^{'}$) is decoded from $frame_{bl}^{en}$ in the lower layer. The quality enhancement layer's data is stored in the residual frame ($frame_{el}^{res}$), which is constructed by original frame ($frame_{bl}$) subtracting the $frame_{bl}^{'}$. Then $frame_{el}^{res}$ is quantized by a QP_{el} which is smaller than QP_{bl} to produce bitstream of the enhancement layer. The amount of information in $frame_{el}^{res}$ is dependent the quantization step sizes as in the spatial scalability. In addition to that, it is also related to the texture. If images of video sequences have more texture (shape or edge) feature, $frame_{el}^{res}$ generally has more texture residual signals. As a result, the quality enhancement layer discloses more semantic information.

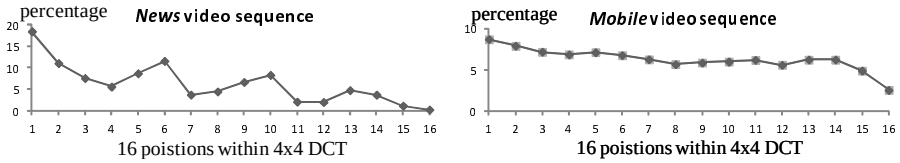
Figure 1(c) and 1(d) illustrate that the quality enhancement layer leaks visual content at different QP sizes, which is more evident after using multimedia tools, such as edge detection and sharpening tools. Entropy statistics of the base layer and quality enhancement layers can be calculated from the number of non-zero coefficients of the base layer and the quality enhancement layer. Note that the non-zero coffficients of the quality enhancement layer is around 1 times more than that of the base layer. In addition, Table 2 measures the leakage by using the ESS scores for six sequences. Images of video sequence such as *Mobile* have richer texture,and have a higher ESS score.

Table 2. ESS scores of $frame_{el}^{'}$ against the $frame_{bl}$ under edge detection

Sequences	*News*	*Football*	*Mobile*	*Soccer*	*Bridge-far*
ESS	0.47	0.45	0.61	0.49	0.54

Media Grain Scalability Similar to CGS, the leakage in MGS is affected by the quantization steps and the richness of texture. Moreover, it is also dependent on the quantization coefficients' distribution which is related to the texture feature of macroblocks.

For the 1st images of *News* and *Mobile* sequences, Figure 5 plots the percentage of non-zero coefficients at 16 positions of 4×4 DCT of them. Figure 5(a) plots a speed gradient for the 1st frame of *News*. It illustrates that more of non-zero quantization coefficients' percentage are below 5% because the macroblocks in *News* are smooth (black background and flat clothes). Figure 5(b) illustrates a relatively flat line for the 1st frame of *Mobile* because it has richer texture so the non-zero quantization coefficients is in a homogeneous distribution. Therefore, if an image has richer texture feature, every MGS enhancement layers can expose visual information.



(a) Smooth texture (b) Richer texture

Fig. 5. MGS quality scalability experiments.

Although both spatial and quality scalability make use of Inter-layer prediction techniques, they leakage contents in different amounts. For quality scalability, $frame_{bl}^{'}$ is typically an effective prediction reference as it is identical to $frame_{bl}$, except for distortion introduced by quantization. Spatial enhancement layers lead to more leakage due to upsampling, which results in more distortion in $frame_{ups}$, as compared in Figure 1(b), Figure 1(c) and Figure 1(d).

Figure 6 plots the binary sizes of frames in both the spatial enhancement layer and the quality enhancement layer, with the same resolution and coding parameter (QP, GOP Size, Intra Interval). This shows that a spatial enhancement frame carries three to six times more information than a quality enhancement frame. Therefore, the former causes more leakage than the latter.

Temporal Scalability The literature [15, 16, 28, 29] have shown that the temporal scalability layer expose content information if not encrypted. For com-

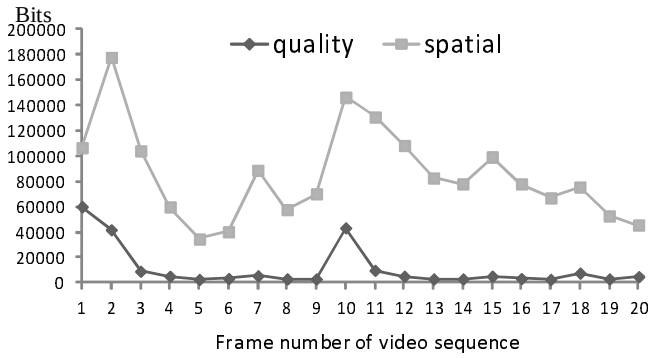


Fig. 6. Bits comparison between quality and spatial scalability.

pleteness of this paper, we briefly review this issue and show our experiment results.

Temporal scalability uses Inter prediction techniques which include selecting a prediction region, generating a prediction block and subtracting this from the original block of samples to form a residual frame $frame_{el}^{res}$. The offset between the object position of the current partition and the prediction region in the reference picture, namely the motion vector, lead to data leakage. Moreover, Intra-coded macroblocks within the temporal enhancement layer (Inter-Frames) is dangerous without encryption. These argument can be verified by measuring the size of the temporal layer frames and the ESS scores.

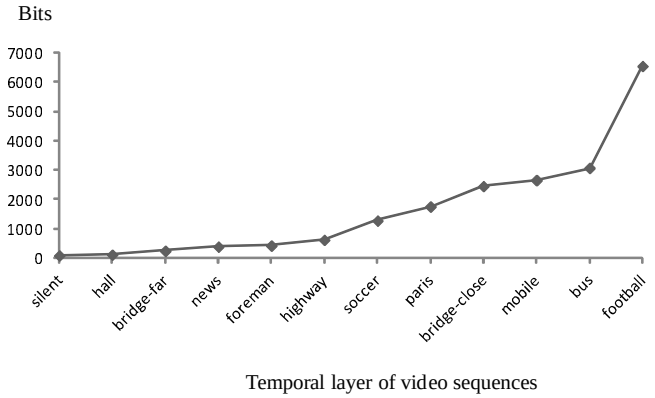


Fig. 7. The relation between motion feature of video sequence and bits at temporal scalability.

Figure 7 shows the bit size of the frame at the third temporal layer for each of the ten video sequences in our test. Static video sequences, (such as *Silent, Hall,*

Bridge-far, News, Foreman, Highway) have fewer bits in the temporal enhancement layer than motion sequences (e.g., *Soccer, Paris, Bridge-close, Mobile, Bus, Football*). Temporal layers of motion video sequences must be encrypted. In addition, Table 3 summarizes the ESS scores of motion and/or Intra-coded areas against their respective areas in the $frame_{el}$, where the Intra-coded block in *Bus* has the highest score.

Table 3. The leakage assessment by ESS and edge detection for temporal scalability

Sequences	*Hall* (human)	*Highway* (line)	*Silent* (hand)	*Foreman* (face contour)	*Bus* (Intra-coded block)
ESS	0.314	0.676	0.368	0.488	1

4.3 Summary

We summarize the leakage of partial encryption in all three scalabilities as below.

- Spatial Scalability: Compared with quality and temporal scalability, in all types of video sequences or the same video sequence with arbitrary QP difference between the reference layer and enhancement layer, the image of spatial scalability leaks more visual content; the larger the QP difference, the more leakage on visual content.
- Quality Scalability: Similar to the spatial scalability, the quality scalability is affected by the QP difference. In addition to that, other factors also lead to the exposure.
  1. Coarse Grain Scalability: The image features, such as texture, shape and edge, lead to content exposure. The more richer the features are, the more data are leaked in the enhancement layers.
  2. Medium Grain Scalability: MGS layers may consist of low frequency, middle frequency, or high frequency DCT coefficients, whose leakage are related to the image features. A rich feature image will contains more non-zero coefficients so that each MGS layer may disclose information.
- Temporal Scalability: The motion feature determines the amount of leakage from the temporal scalability layers. Obviously, more intensive motions result in more leakage in the enhancement layers.

5 Related Work

In [17], Yu gave an overview of scalable encryption schemes which summarize previous works on selective encryption, format compliant encryption, and progressive encryption on scalable multimedia. At the same time, the article concluded that only part of entire bitstream are encrypted while the rest are left

in the clear. Further, Yu addressed the advantages of scalable encryption for wireless multimedia communication and presented improvement on scalability via progressive encryption.

Zhu et al. described in [15] the general requirements and desirable features of an encryption system for scalable multimedia, such as encrypted content leakage (perceptibility), security and scalability, and presented a survey of the current state of the art of technologies in scalable encryption and analyzed the performances (leakage, overhead and complexity) of encryption schemes for scalable multimedia (JPEG2000 and MPEG-4 FGS). The article concluded that naive encryption algorithm is inappropriate for encryption of scalable code stream because scalability is completely removed. Meanwhile, after reviewing selective encryption on JPEG2000 and MPEG-4 FGS, the authors pointed out selective encryption usually leaks some information of the encrypted content and is less secure, and encryption of the base layer of MPEG-4 FGS alone may not acceptable in some applications.

In [30], it gave a brief overview of the concept, desirable feature and possible attacks on multimedia encryption. Before the description of prototype for multimedia encryption, the article introduced the symmetric key encryption (block and stream cipher) and cryptanalysis. In addition, desirable requirement, characteristics and attacks of multimedia encryption were discussed in this article. During introduction of multimedia encryption, the authors classed them with total encryption, selective encryption, perceptual encryption, joint compression encryption, format compliant encryption, and scalable encryption. For scalable encryption, it reviewed various scalable encryption techniques of JPEG2000 and MPEG-4 FGS and showed that some of them have problem of content leakage if selective encryption is given.

Lian [18] described the partial encryption in which their performances, such as security, encryption efficiency, compression efficiency, and format compliance were analyzed and compared in chapter 5, and showed that some partial encryption schemes were not secure enough due to partitioning and part selection. Meanwhile, chapter 8 classified scalable encryption with layered encryption, layered and progressive encryption, progressive encryption and scalable encryption according to the scalable property.

6 Conclusion

In this article, we investigated whether partial encryption in H.264/SVC can protect data confidentiality. Our experiments showed that unencrypted enhancement layers leak significant context information about the video stream, from all three scalability dimensions. We also analyzed the coding techniques for spatial, quality and temporal scalabilities, and showed that the coding techniques used in H.264/SVC determine that enhancement layers have to be encrypted for the confidentiality purpose, although partial encryption may be sufficient for access control in the sense of deterring unauthorized access to the complete high quality video.

Acknowledgment

This work was supported by A*STAR SERC Grant No. 102 101 0027 in Singapore. The authors would like to thank Mr. Yifan Zhao for his help in conducting some of the experiments.

References

1. W. Li. Overview of Fine Granularity Scalability in MPEG-4 Video Standard. *IEEE Transactions on Circuits and System for video Technology*, 11(3): 301-317, 2001.
2. H. Schwarz, D. Marpe, and T. Wiegand. Overview of the scalable video coding extension of the h.264/avc standard. *IEEE Transactions on Circuits and System for Video Technology*, 17(9):1103-1120, September 2007.
3. M. Wien, H. Schwarz, and T. Oelbaum. Performance Analysis of SVC. *IEEE Transactions on Circuits and System for Video Technology*, 17(9):1194-1203, September 2007.
4. H. Sohn, W. De Neve and Y. M. Ro. Region-of-interest scrambling for scalable surveillance video using JPEG XR. *ACM Multimedia*, page 861-864, 2009.
5. H. Sohn, W. De Neve and Y. M. Ro. Privacy Protection in Video Surveillance Systems: Analysis of Subband-Adaptive Scrambling in JPEG XR. *IEEE Trans. Circuits Syst. Video Techn*, 21(2): 170-177, 2011.
6. C. Yuan, B.B. Zhu, Y. Wang, S. Li, and Y. Zhong. Efficient and fully scalable encryption for MPEG-4 FGS. *ISCAS (2)*, page 620-623, 2003.
7. B.B. Zhu, C. Yuan, Y. Wang, and S. Li. Scalable protection for MPEG-4 fine granularity scalability. *IEEE Transactions on Multimedia*, 7(2): 222-233, 2005.
8. Z. Shahid, M. Chaumont and W. Puech. Selective and scalable encryption of enhancement layers for dyadic scalable H.264/AVC by scrambling of scan patterns. *ICIP*, page 1273-1276, 2009.
9. G. Algin and E. Tunali. Scalable video encryption of H.264 SVC Codec. *J. Visual Communication and Image Representation (JVCIR)*, 22(4): 353-364, 2011.
10. A. Tosun and W. Feng. Efficient Multi-layer Coding and Encryption of MPEG Video Streams. *IEEE International Conference on Multimedia and Expo (I)*, page 119-122, 2000.
11. C. Li, X. Zhou, and Y. Zhong. Nal Level Encryption for Scalable Video Coding. *PCM*, pages 496-505, 2008.
12. C. Li, C. Yuan and Y. Zhong. Layered Encryption for Scalable Video Coding. *2nd International Congress on Image and Signal Processing*, pages 1-4, 2009.
13. W. Zeng and S. Lei. Efficient Frequency Domain Selective Scrambling of Digital Video. *IEEE Transactions on Multimedia*, 5(1): 118-129, 2003.
14. F. Liu and H. Koenig. A Survey of Video Encrytpion Algorithms. *Computer & Security*, 29(1): 3-15, 2010.
15. B. Zhu, M. Swanson and S. Li. Encryption and Authentication for Scalable Multimedia: Current State of the Art and Chanllenges. *Internet Multimedia Management Systems V*, pages 157-170, 2004.
16. X. Liu and A. Eskicioglu. Selective Encryption of Multimedia Content in Distribution Network: Chanllenges and New Directions. *2nd IASTED Int. Conf. on Comm., Internet, and Info. Technol*, pages 527-533, 2003.
17. H. Yu. An Overview on Scalable Encrytpion for Wireless Multimedia Access. *Internet Quality of Service*, pages 24-34, 2003.

18. S. Lian. Multimedia Content Encryption: Techniques and Applications. *CRC Press*, 29(1): 121-130, 2008.
19. T. Stütz and A. Uhl. "A survey of h.264 avc/svc encryption," *IEEE Trans. Circuits Syst. Video Techn.*, 22(3): 325–339, 2012.
20. R.L. Rivest. "The RC4 Encryption Algorithm" *RSA Data Security, Inc.*, March 12, 1992.
21. S.-W. Park and S.-U. Shin. Combined scheme of encryption and watermarking in h.264/scalable video coding (svc). *New Directions in Intelligent Interactive Multimedia*, pages 351-361, 2008.
22. S.-W. Park and S.-U. Shin. Efficient selective encryption scheme for the h.264/scalable video coding(svc). *NCM*, pages 371-376, 2008.
23. S.-W. Park and S.-U. Shin. An efficient encryption and key management scheme for layered access control of h.264/scalable video coding. *IEICE Transactions (IEICET)*, 92-D(5):851-858, 2009.
24. G. Hong, c. Yuan, Y. Wang, and Y. Zhong. A Quality-Controllable Encryption for H.264/AVC Video Coding. *PCM*, page 510-517, 2006.
25. Joint scalable video model software, http://ip.hhi.de/imagecom g1/savce/downloads/svc-reference-software.htm.
26. K. Engel, M. Hadwiger and J. M. Kniss, C. Rezk-Salama, and D. Weiskopf. Real-time volume graphics. *A K Peters*, pages I-XVII, 1-497, 2006.
27. Y. Mao and M. Wu. Security evaluation for communication-friendly encryption of multimedia. *Proceedings of the IEEE International Conference on Image Processing*, 2004.
28. T. Kunkelmann and U. Horn. Video Encryption Based on Data Partitioning and Scalable Coding - A Comparison. *IDMS*, pages 95-106, 1998.
29. I. Agi and L. Gong. An Empirical Study of Secure MPEG Video Transmissions. *ISOC-SNDSS*, pages 137-144, 1996.
30. N. Kulkarni, B. Raman and I. Gupta. Multimedia Encryption: A Brief Overview. *Recent Advances in Multimedia Signal Processing and Communications*, pages 417C449, 2009.