Format-compliant Encryption of H.264/AVC and SVC

Thomas Stütz and Andreas Uhl
University of Salzburg
Department of Computer Sciences
Jakob-Haringerstr. 2
5020 Salzburg, Austria
{tstuetz,uhl}@cosy.sbg.ac.at

Abstract

An encryption approach for H.264/AVC and SVC is proposed. Although the bitstream (format stream) is encrypted with state-of-the-art symmetric ciphers, H.264/AVC and SVC compliance is preserved. Standard compliant encoder/decoder and conventional symmetric ciphers, e.g., in specialized hardware, can still be employed – a significant advantage compared to previous work. The approach is suitable for a wide range of application scenarios.

1. Introduction

The protection of multimedia data has become increasingly important in recent years. There are many real-world application scenarios that require that multimedia data, i.e., visual data / video, are safely protected (see sec. 3.1): digital content distribution over insecure channels, video streaming, video conferencing, digital TV broadcasting, pay-per-view TV, the protection of surveillance video data, and secure erosive storage, e.g., for medical applications.
The most secure method for the protection of multimedia data, sometimes referred to as the naive method, is to encrypt the entire multimedia data with the aid of a cryptographically strong cipher like AES. Unfortunately, the naive method is not able to meet requirements imposed by real-world application scenarios (see sec. 3).
Multimedia data are usually given in a specific format, the most recent video compression standard is H.264/AVC which is both an officially published ISO standard (ISO/IEC 14496-10, also referred to as MPEG-4 Part 10, Advanced Video Coding) and an ITU-T recommendation (ITU-T Rec. H.264). In this paper the term SVC is exclusively used to denote the scalable extension of H.264/AVC, which is specified in Annex G [6], and the term H.264/AVC is used to denote the original, non-scalable format (i.e., format streams conforming to a profile and level as specified in Annex A [6]), while H.264/AVC/SVC refers to the entire standard and the formats defined therein, especially including Amd. 3 Scalable video coding.
SVC has been designed to satisfy the requirements of modern video transmission and storage systems, which are characterized by a wide range of connection quality and receiving devices. The scalability of the SVC format allows easy rate adaptation in any of the scalable dimensions (temporal, spatial, quality) in the compressed domain by removal of parts of the SVC format stream. Both H.264/AVC and SVC are discussed in more detail in section 2.
Standardized multimedia formats are of overwhelming importance, as they guarantee the interoperability of various software and hardware systems. The necessity of developing security tools for specific multimedia data formats has found its implementation for JPEG2000 in Part 8 of the JPEG2000 standard suite [5]. In this work we propose an approach that elegantly and seamlessly integrates security (encryption) tools in the H.264/AVC/SVC standard. In it the syntax and semantics shared by both data formats are employed. The integration of security (encryption) tools is explained in section 4 in more detail.
Based on the requirements and evaluation criteria developed in section 3 our approach is evaluated and compared to previously presented approaches, specifically for H.264/AVC, SVC and JPEG2000 (see sec. 5). Finally, in section 6 we draw our conclusions.

2. Overview of H.264/AVC and SVC

Several years have passed since the first publication of the H.264/AVC standard in 2003 and only recently amendment 3 (SVC) has been published [6]. Note that in the H.264/AVC/SVC standard a stream satisfying any of the therein specified formats is denoted bitstream, which we prefer to denote format stream throughout this work to point out that the corresponding stream is not just an arbitrary bitstream, but follows the H.264/AVC/SVC format syntax and

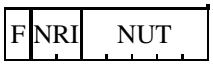

Figure 1. NAL unit header structure.

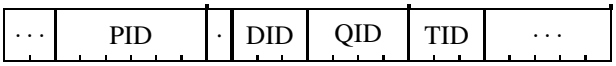

Figure 2. NAL unit header SVC extension structure.

semantics. The basic design of H.264/AVC is similar to previous video coding standards, but the numerous differences and new features result in a significantly improved compression performance [11]. The basic steps are subdivision of the picture into macroblocks (16x16 blocks of luma samples, which may be further subdivided down to 4x4 blocks), inter or intra prediction, transformation and quantization, entropy coding, and NAL unit assembly.

A major design requirement for SVC has been the backwards compatibility to the existing H.264/AVC standard. I.e., SVC format streams are valid H.264/AVC format streams (format-compliant with respect to the non-scalable H.264/AVC format [6, sec. C.3]) and thus decodeable by H.264/AVC compliant decoders [6, sec. C.4]. Major parts of the H.264/AVC video coding system have been adopted. An SVC format stream contains a base layer and one or more enhancement layers each may augment the user experience in one of three dimensions (temporal/spatial/quality). A format stream is temporally scalable if it contains substreams with lower frame rates. Due to the flexible inter prediction in H.264/AVC, the implementation of temporal scalability within H.264/AVC/SVC has been straightforward by employing special prediction structures, e.g., dyadic temporal enhancement layers with hierarchical B-pictures. But temporal scalability in H.264/AVC/SVC is not limited to dyadic prediction structures. A format stream is spatially scalable if it contains substreams with different resolutions. Spatial scalability with arbitrary resolutions is supported. A format stream is quality scalable if it contains substreams with different qualities (in a SNR sense), but same resolution. SVC is capable of offering quality scalability.

2.1. H.264/AVC/SVC: The Network Abstraction Layer

One of the fundamental issues in the developing of SVC was its integration into the existing H.264/AVC standard. Largely responsible for the successful integration was the conceptually clear structure of H.264/AVC, which distinguishes between a coding layer (VCL, video coding layer, and non-VCL, non video coding layer) and a network abstraction layer (NAL). The VCL is responsible for creating a coded representation of the moving pictures, while the NAL formats these data and provides header information in a simple and effective fashion, i.e., a NAL unit header is not entropy coded. VCL data are organized into NAL units, which start with a one byte header. Most important is the type of a NAL unit (NUT) that is inferred from the NAL unit header (its structure is illustrated in fig. 1). The F (forbidden_zero_bit) shall always be equal to 0, the semantics of the value of NRI (nal_ref_idc) relate to relative importance of the corresponding NAL unit. Depending on the NUT (NAL unit type) the subsequent data are interpreted. Most importantly, H.264/AVC only specifies semantics for a subset of the 32 possible values of the NUT and concisely specifies a H.264/AVC compliant decoder's behavior if a NAL unit with a reserved or unspecified NUT is encountered. Decoders shall ignore (remove and discard from the format stream) the contents of all NAL units that use a reserved or unspecified value of the NUT. There is a significant semantic difference between reserved and unspecified NUT types. Unspecified NUTs may be used without any restriction, as they will never be assigned any normative meaning in the H.264/AVC/SVC standard. Thus applications are free to use those values for the NUT for their specific needs and requirements. The resulting format stream including these NAL units is still format-compliant with respect to H.264/AVC/SVC. Reserved NUT values are different, because future amendments to the H.264/AVC/SVC standard may assign specific semantics and corresponding decoding processes for their subsequent data. Nonetheless, a current H.264/AVC/SVC compliant decoder has to ignore NAL units with a reserved NUT value and the format stream containing such NAL units is still format-compliant with respect to H.264/AVC/SVC, but may violate the syntax and semantics of future extensions.

In table 1 the most important NUTs for the scope of this work are summarized and additionally their classification into either VCL or non-VCL is given for both H.264/AVC

NUT	Description	AVC class	SVC class
0	Unspecified	Non-VCL	Non-VCL
1	Non-IDR slice	VCL	VCL
5	IDR slice	VCL	VCL
6	SEI	Non-VCL	Non-VCL
12	Filler data	Non-VCL	Non-VCL
14	Prefix NAL	Non-VCL	Variable
16 . . . 18	Reserved	Non-VCL	Non-VCL
20	SVC slice	Non-VCL	VCL
21 . . . 23	Reserved	Non-VCL	Non-VCL
24 . . . 31	Unspecified	Non-VCL	Non-VCL

Table 1. Selected NAL unit types.

(in the column labelled AVC) and SVC. In an IDR-slice only intraprediction is applied.
Enhancement layer data are contained in NAL units with a previously reserved NUT (14 and 20), thus these data are simply ignored by a H.264/AVC compliant decoder and the remaining part of the format stream is interpreted as a valid and decodeable H.264/AVC format stream. An SVC NAL unit header extension is specified (see fig. 2), which contains valuable information about the NAL unit content, such as the PID (priority id), the DID (dependency id), the QID (quality id) and the TID (temporal id).

3. Multimedia Encryption

Numerous contributions to the field of multimedia encryption have been published in recent years and the interested reader is referred to [14] and [2] for extensive overviews.

3.1. Application Scenarios

For many application scenarios the naive method, i.e., the encryption of the multimedia data with a cryptographic cipher, is not suitable, as its application in a rate adaptation scenario enforces a compromise of security or an increase of computational complexity (see sec. 3.1.1), or a dissipation of bandwidth or storage capacity (see sec. 3.1.2).

3.1.1 Secure Adaptation

One benefit of SVC is that rate adaptions can be conducted easily by simply dropping parts of the format stream, i.e., certain NAL units. In a streaming scenario, the rate adaption will be conducted by MANEs (Media-Aware Network Elements). Unfortunately, if the naive method is applied, the structure of the SVC format stream is lost and thus the ability to perform efficient rate adaption [1]. As a last resort, the MANE has to decrypt the efstream (encrypted format stream) and adapt the rate and re-encrypt the adapted format stream. Thus the MANE must have access to the secret encryption key, which severely compromises the security, as the key has to be transmitted to an untrusted or insecure third party, i.e., the MANE.
Real-world applications of secure adaptation in a network scenario are, e.g., video conferencing, digital TV broadcasting, pay-per-view TV, and the protection of surveillance video data. The preservation of scalability in the encryption process is fundamental for secure erosive storage, i.e., visual data are compressed and encrypted once and high quality parts of the efstream are deleted after some expiration time.

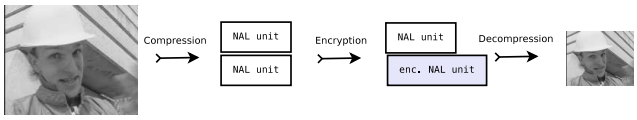

Figure 3. Conditional access.

3.1.2 Perceptual / transparent encryption

Compared to conventional encryption, perceptual / transparent encryption has an entirely different goal: the visual data has to be discernible after encryption, but severely degraded (depending on a perceptual quality factor). For scalable formats, perceptual / transparent encryption can be realized with conditional access techniques. If certain consumers only have restricted access to certain parts of the format stream, e.g., only certain pictures of a video sequence, this is referred to as conditional access (see fig. 3 for an example; only a low resolution video is decodeable). Real-world applications are mainly found in the commercial area, such as digital TV broadcasting and pay-per-view TV, where the broadcast data shall also attract possible customers (currently a frequently applied practice of pay TV broadcasters).

3.2. Classification

On the basis of previous work [14, 2, 13], a classification of multimedia encryption is discussed in this section. A classification can be carried out on the basis of the properties of the efstream. Important properties of the efstream are:

E1 Format-compliance

E2 Scalability

E3 Compression-equivalence

E4 Perceptual quality of the decompressed efstream (decodability as ensured by format-compliance is a prerequisite)

The format-compliance (E1) and the perceptual quality (E4) of an efstream are of fundamental importance for perceptual / transparent encryption (see sec. 3.1.2), while the preservation of scalability (E2) is essential for secure adaptation (see sec. 3.1.1). Only a negligible increase of the size of the efstream, compared to the original format stream, is desired (E3).
A further classification can be conducted on the basis of technical and functional properties of the multimedia encryption approach.

T1 Selective / partial application

T2 Conventional / lightweight encryption algorithms

T3 Encryption before / after / during compression

T4 Application of standard compliant encoders and decoders

A multimedia encryption approach is called selective / partial if only a subset of format stream data is encrypted (T1). In order to lower computational demands the application of less secure (lightweight / soft) encryption algorithms has been proposed (T2). The doubtful security of such algorithms is an argument against their application. It is of fundamental importance at which stage in the compression pipeline encryption is conducted (T3). In [14] it is pointed out that the complexity of encryption compared to the complexity of compression is negligible (discussed in detail for JPEG2000 and AES). Therefore, the encryption after compression is preferable for many application scenarios, e.g., it is unfeasible in the case of pay-per-view TV to apply compression-integrated encryption for every customer, as the computational complexity will be substantially too high. Another drawback of compression-integrated approaches (i.e., approaches that conduct encryption during compression) is that the encoder and the decoder have to be modified, which rules out the application of standard compliant compression hardware (which can not be modified easily, cf. T4).
In a certain application scenario, one can assess the following properties of a multimedia encryption approach:

A1 Complexity

A2 Compression

A3 Security

While complexity and compression performance can be simply measured, the definite assessment of the security of multimedia encryption approach is tedious and troublesome. In [12] the authors discuss the security of selective / partial encryption schemes and propose a model for their respective analysis.

4. Format-compliant Encryption of H.264/AVC/SVC

As discussed in section 2.1, the NAL unit data is processed as determined by the value of NUT (NAL unit type) in the NAL unit header. If an unspecified NUT is encountered, the corresponding NAL unit has to be ignored by a standard-compliant decoder (the format stream containing NAL units with an unspecified NUT is still format-compliant). We propose to set NUT of NAL units selected for encryption to an unspecified value and encrypt the NAL unit payload. In order to preserve the compression performance of H.264/AVC/SVC, a distinct unspecified NUT value is selected for very frequent NUT values (see table 2). In the case of SVC NAL units (NUT 14 and 20) the SVC NAL header extension (3 byte) is preserved and only the NAL unit payload is encrypted. For less frequent NUT values, a common NUT value (27) is used and the header byte of the plaintext NAL unit is the first payload byte, which is followed by the the encrypted NAL unit payload.
In order to preserve the full scalability of the format stream, NAL units have to be independently decryptable.
A straightforward method is to use a state-the-art block cipher (we apply AES) in ECB (Electronic Codebook Mode) to encrypt the NAL unit payload. If the NAL unit payload is not a multiple of the block size of the cipher, then ciphertext stealing is applied. Let us summarize the steps of our proposed format-compliant encryption approach:

1. The raw video is compressed with H.264/AVC/SVC (for many application scenarios the data is already available in the compressed format).

2. According to the desired application scenario, NAL units are selected for encryption (see sec. 5.2).

3. The selected NAL units are processed depending on the value of the NUT:
   NUT 1 and 5: For NAL units with a NUT of 1 the NUT is set to 0 and for NAL units with a NUT 5 to 24, the remaining header fields, namely F and NRI (see sec. 2.1), are preserved. The NAL unit payload is encrypted.
   NUT 14 and 20: For SVC NAL unit the encryption process is slightly different. For NAL units with a NUT of 14 the NUT is set to 25 and for NAL units with a NUT of 20 to 26. The remaining header fields are preserved as well as the SVC NAL unit header extension (see sec. 2.1). The remaining NAL unit payload is encrypted.
   All other NUT values: A new NAL unit header is constructed with a NUT of 27, the remaining header fields are set to values of the NAL unit to be encrypted. The original NAL unit header is the first byte of the new NAL unit payload. The plaintext header is followed by the encrypted original NAL unit payload.
   The efstream (encrypted format stream) is transmitted (rate adaptation may be conducted).

It has to be pointed out that only selections of NAL units that preserve valid SVC substreams result in format-compliant efstreams.
Note that this approach could also be integrated “officially” as part of a new security extension into H.264/AVC/SVC, only that reserved NUTs instead of unspecified NUTs

NUT	new NUT
1	0
5	24
14	25
20	26
other	27

Table 2. New NUTs for encrypted NAL units.

would be used to that end. Alternatively to a block cipher in ECB mode a block cipher in OFB-mode or counter mode could be used with a unique IV (initialization vector) that is individually constructed for each NAL unit. The construction of a unique IV for every NAL unit, which is robust to any valid format stream adaptation (by discarding NAL units), is not trivial.

5. Evaluation and Comparison

First we will analyze our proposed format-compliant encryption scheme on the basis of the presented classification, then we will discuss its suitability for the discussed application scenarios and finally we compare our approach to previous work.

5.1. Properties

Based on the classification for multimedia encryption schemes presented in section 3.2, we analyze our proposed encryption approach. The produced efstream is format compliant (E1), scalable (E2) (if the plaintext format stream is scalable of course), has the same size, except for a negligible number of bytes (E3), and the reconstructible perceptual quality can be adjusted to any quality contained in the original format stream (E4). Thus in conjunction with SVC any desired quality can be achieved, simply by selecting the NAL units of the enhancement layers for encryption (see sec. 3.1.2).
Partial / selective application is possible (T1), i.e., only a subset of all NAL units is encrypted. This is the case for the implementation of perceptual / transparent encryption with our approach. Conventional encryption algorithms are applied (T2) and thus the security of the encrypted data is not subject to discussion. This property ensures also that highly optimized software (e.g., AES implementations) and specialized hardware (e.g., AES chips) can be employed for encryption, a potentially significant advantage. The encryption is applied after compression (T3) and is therefore not bounded to computationally complex compression. Standard compliant H.264/AVC/SVC encoders and decoders are employed (T4), which again allows the application of highly optimized software and special hardware for H.264/AVC/SVC compression.
In general (for all application scenarios) the complexity of the encryption approach is very low (A1), the compression is not affected (A2) and the security can be considered high (A3).

5.2. Suitability for Application Scenarios

The proposed format-compliant encryption approach can satisfy all of the discussed application scenarios in conjunction with SVC (see sec. 3.1). For secure adaptation of an SVC format stream, all VCL NAL units are encrypted (according to the SVC class). As the SVC header extension is preserved, so is the scalability.
Conditional access can be implemented as well on the granularity of NAL units and thus perceptual / transparent encryption can be implemented with SVC. We assume that the target quality (i.e., the public low quality version of a video) is contained in the base layer. To that end all VCL NAL units that are not part of the base layer are selected for encryption, only preserving the low quality base layer. The resulting efstream may be distributed freely and those interested in the high quality of the content can buy the secret key necessary for decryption of the enhancement quality NAL units.

5.3. Comparison with prior work

The novelty of this approach is to directly integrate encryption in the H.264/AVC/SVC syntax. Thus we can preserve format-compliance, although our approach is applied after the compression. This means a significant advantage as both the preservation of format-compliance and the separation of encryption from compression are of considerable importance in order to be able to implement certain application scenarios and keep the computational complexity low. As the complexity of encryption compared to compression is marginal, the overall reduction of complexity in application scenarios, that require the encryption of the visual data with many different keys is enormous, e.g., in a pay-per-view TV application scenario every customer has a separate private key.
Format-compliance and post-compression application of encryption are often even regarded as contradicting, e.g., in [13] it is stated that "Post-compression approaches are inherently format-defiant" (format-defiant denotes not format-compliant). For H.264/AVC/SVC the proposed approach (see sec. 4) is the first to offer both. Almost all of the previously proposed approaches implement encryption during compression, e.g., the scrambling of the intra prediction modes [3] or of motion vector data [8], the encryption of

coefficient data and the perturbation of motion vectors [9], and the encryption of coefficient signs [10].

Therefore the computationally demanding compression has to be conducted for encryption and decryption, a drawback our approach does not have. However, our approach is not suitable for perceptual / transparent encryption for plain H.264/AVC.

There are proposals on the basis of MPEG-21 [4] that rely on the application of MPEG-21; an assumption that may not hold. Given the numerous container and meta formats for video data, the integration of security tools (e.g., encryption) within the video codec solves the problem once and security tools do not have to be integrated in every container and meta format.

Previous work on SVC is primarily based on a draft standard that has significantly changed (e.g., FGS has been removed). In [1] principles for secure scalable streaming (basically secure adaptation in a network) and SVC are discussed. In [15] sign encryption of "texture, motion vector, and FGS data" is proposed, and in [7] this idea is extended to protect regions of interest. In contrast to our approach, the approaches of [15, 7] are compression-integrated and in contrast to [1] a concrete (and implementable) encryption approach is given.

There have been numerous format-compliant encryption proposals for JPEG2000 [16]. These encryption approaches require specifically designed encryption algorithms in order to preserve the format-compliance. None has been integrated into the normative tools of the JPSEC standard [5].

6. Conclusion

A format-compliant encryption approach specific to H.264/AVC/SVC has been proposed, evaluated, and compared to prior work. The novelty of this approach is the seamless integration into H.264/AVC/SVC, which utilizes the syntax and semantics of the NAL (network abstraction layer) to that end. The approach preserves the scalability of SVC and is therefore applicable in advanced application scenarios, such as secure adaptation. As format-compliance is preserved, perceptual / transparent encryption can also be implemented with the proposed approach; most interesting for commercial applications. Compression and encryption are not interleaved, allowing efficient encryption with diverse keys.

References

[1] J. Apostolopoulos. Architectural principles for secure streaming & secure adaptation in the developing scalable video coding (SVC) standard. In *Proceedings of the IEEE International Conference on Image Processing, ICIP '06*, pages 729–732, Oct. 2006.

[2] B. Furht, E. Muharemagic, and D. Socek. *Multimedia Encryption and Watermarking*, volume 28 of *Multimedia Systems and Applications*. Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 2005.

[3] G.-M. Hong, C. Yuan, Y. Wang, and Y. Zhong. A quality-controllable encryption for H.264/AVC video coding. In Y. Zhuang, S. Yang, Y. Rui, and Q. He, editors, *Proceedings of the 7th Pacific Rim Conference on Multimedia, Advances in Multimedia Information Processing, PCM'06*, volume 4261 of *Lecture Notes in Computer Science*, pages 510–517. Springer-Verlag, Nov. 2006.

[4] R. Iqbal, S. Shirmohammadi, A. E. Saddik, and J. Zhao. Compressed-domain video processing for adaptation, encryption, and authentication. *IEEE Multimedia*, 15(2):38–50, Apr. 2008.

[5] ISO/IEC 15444-8. Information technology – JPEG2000 image coding system, Part 8: Secure JPEG2000, Apr. 2007.

[6] ITU-T H.264. Advanced video coding for generic audivisual services, Nov. 2007.

[7] Y. Kim, S. Yin, T. Bae, and Y. Ro. A selective video encryption for the region of interest in scalable video coding. In *Proceedings of the TENCON 2007 - IEEE Region 10 Conference*, pages 1–4, Taipei, Taiwan, Oct. 2007.

[8] S. G. Kwon, W. I. Choi, and B. Jeon. Digital video scrambling using motion vector and slice relocation. In *Proceedings of Second International Conference of Image Analysis and Recognition, ICIAR'05*, volume 3656 of *Lecture Notes in Computer Science*, pages 207–214, Toronto, Canada, Sept. 2005. Springer-Verlag.

[9] E. Magli, M. Grangetto, and G. Olmo. Conditional access to H.264/AVC video with drift control. In *Proceedings of the IEEE International Conference on Multimedia and Expo, ICME'06*. IEEE, July 2006.

[10] T. Nithin, D. Lefol, D. Bull, and D. Redmil. A novel secure H.264 transcoder using selective encryption. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'07)*. IEEE, Sept. 2007.

[11] I. E. G. Richardson. *H.264 and MPEG-4 video compression: video coding for next generation multimedia*. Wiley & Sons, 2003.

[12] A. Said. Measuring the strength of partial encryption schemes. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'05)*, volume 2, Sept. 2005.

[13] D. Socek, H. Kalva, S. Magliveras, O. Marques, D. Ćulibrk, and B. Furht. New approaches to encryption and steganography for digital videos. *Multimedia Systems*, 13(3):191–204, 2007.

[14] A. Uhl and A. Pommer. *Image and Video Encryption. From Digital Rights Management to Secured Personal Communication*, volume 15 of *Advances in Information Security*. Springer-Verlag, 2005.

[15] 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*, volume 4283 of *Lecture Notes in Computer Science*, pages 407–421, Korea, Nov. 2006. Springer.

[16] Y. Yang, B. B. Zhu, Y. Yang, S. Li, and N. Yu. Efficient and syntax-compliant JPEG2000 encryption preserving original fine granularity of scalability. *EURASIP Journal on Information Security*, 2007.