A Survey of H.264 AVC/SVC Encryption- Part I

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 22, NO. 3, MARCH 2012

- **This Talk intends to give researchers and practitioners an analytic and critical overview of the state-of-the-art of video encryption narrowed down to its joint application with the H.264 standard suite and associated protocols (packaging/ streaming) and processes (transcoding/watermarking).**

I. Introduction

- H.264 is the most widely-deployed video compression system and has gained a dominance comparable only to JPEG for image compression. The H.264 standard has also been extended to allow scalable video coding (referred to as SVC within this talk) with a backwards compatible non-scalable base layer (referred to as AVC in this talk).

- This extension enables the implementation of advanced application scenarios with H.264, such as scalable streaming and universal multimedia access.
- Given the dominant application of H.264 as video compression system, the necessity of practical security tools for H.264 is unquestionable.

- A secure approach to encrypt H.264, also referred to as “naive” encryption approach, is to encrypt the entire compressed H.264 bitstream with a secure cipher, e.g., AES, in a secure mode, e.g., cipher block chaining (CBC) mode.

- There are well-founded reasons not to stick to this approach, but to apply specifically designed encryption routines.

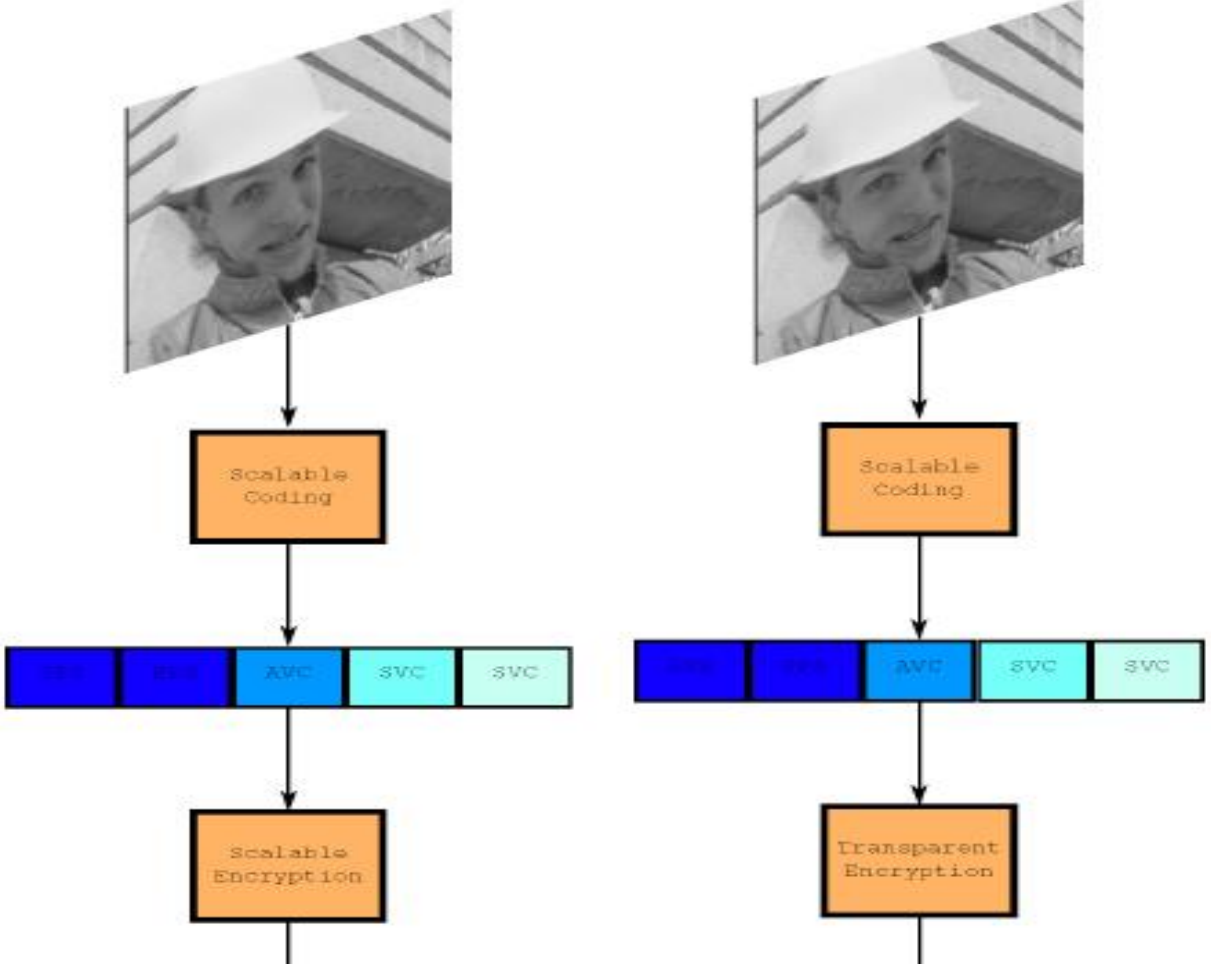
- 1) The implementation of advanced application scenarios, such as secure adaptation, transparent/perceptual encryption and privacy preserving encryption.

- 2) The preservation of properties and functionalities of the bitstream, such as format-compliance, scalability, streaming/packetization, fast forward, extraction of subsequences, transcodability, watermarking, and error resilience.
- 3) The reduction of computational complexity (especially in the context of mobile computing).

- Secure adaptation requires a scalable bitstream and specific encryption routines that preserve the scalability in the encrypted domain [see Fig. 1(a)].
- Secure adaptation is the basis for secure scalable streaming, where secure adaptation is employed in a multimedia streaming scenario.

- A secure stream for a mobile phone (low bandwidth, low resolution display, low computing power) and a personal computer (high bandwidth, high resolution display, high computing power) can be generated from the same secure source stream (by secure adaptation) without the necessity of the secret key, thus enabling creator-to-consumer security.

- Transparent encryption denotes encryption schemes where a low quality can be decoded from the ciphertext; this functionality can be implemented with scalable bitstreams [see Fig. 1(b)] by encryption of the enhancement layers.
- Privacy preserving encryption should conceal the identity of persons, an exemplary implementation is shown in Fig. 2.


Scalable Coding
AVC
SVC
SVC
Scalable Encryption
Scalable Coding
AVC
SVC
SVC
Transparent Encryption

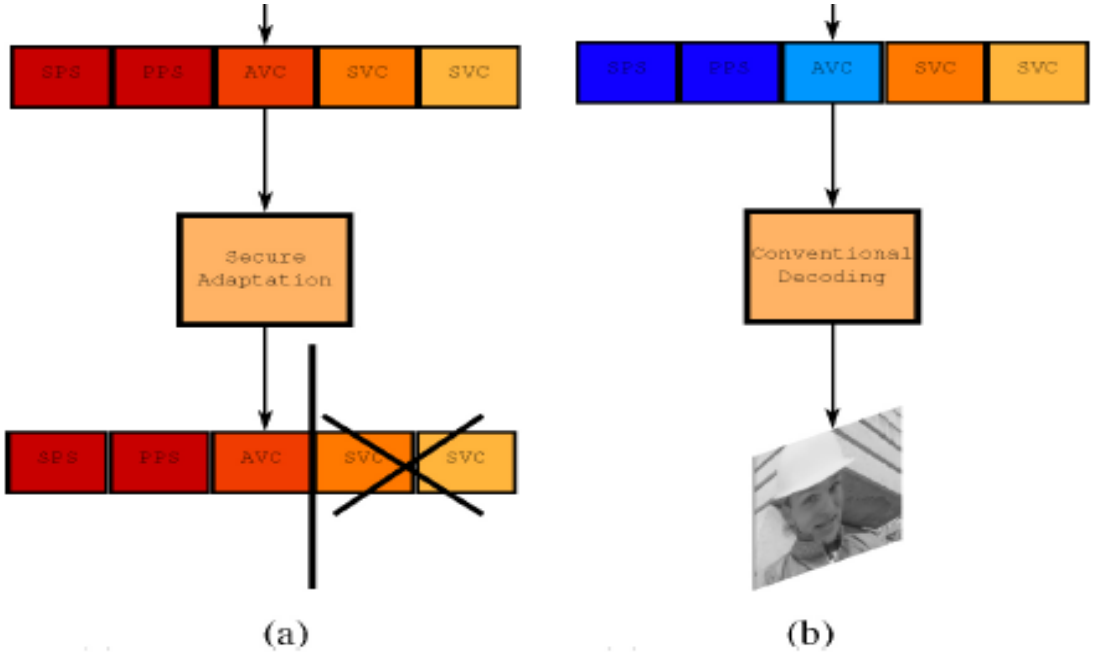


Fig. 1. Secure adaptation and transparent encryption. (a) Secure adaptation. (b) Transparent encryption.

Fig. 2. Privacy preserving encryption: DCT coefficients permutation (figure taken from [13, Fig. 2(b), p. 1171]).

II. Overview of H.264 AVC/SVC

- The H.264 standard specifies the syntax and semantics of the bitstream together with a normative decoding process [1].
- However, it is often and especially in the context of H.264 encryption more convenient to consider the encoding process.

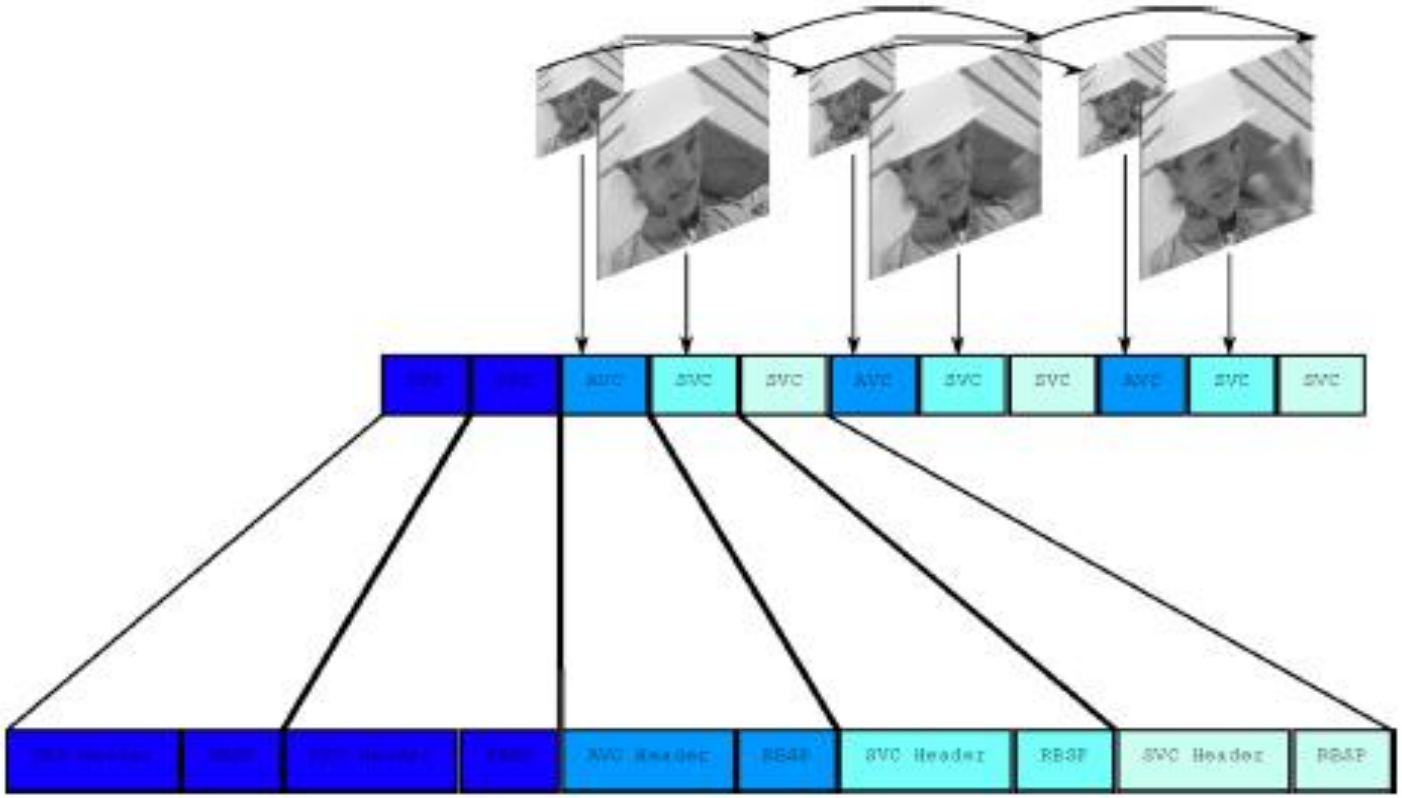


Fig. 3. Mapping of video data to H.264 SVC NALUs.

- A typical H.264 encoder has the structure as outlined in Fig. 4. Important parts are motion estimation (ME in Fig. 4) and motion compensation (MC in Fig. 4). Novelties in H.264 compared to previous video coding standards are intraprediction (Intra pred in Fig. 4) and in-loop deblocking filtering, i.e., reference pictures are filtered to reduce blocking artifacts prior to motion estimation and compensation.

- A 4×4 DCT-based transform is applied (T in Fig. 4), followed by quantization (Q in Fig. 4). There are two types of entropy encoding in H.264, namely context adaptive variable length coding (CAVLC) and context adaptive binary arithmetic coding (CABAC).

- The scalable extension of H.264, referred to as SVC, employs most of the tools defined in the nonscalable H.264, referred to as AVC. An SVC bitstream consists of a base layer and enhancement layers; each enhancement layer improves the video in one of three "scalability dimensions," namely resolution (spatial), quality (SNR), and time (temporal).

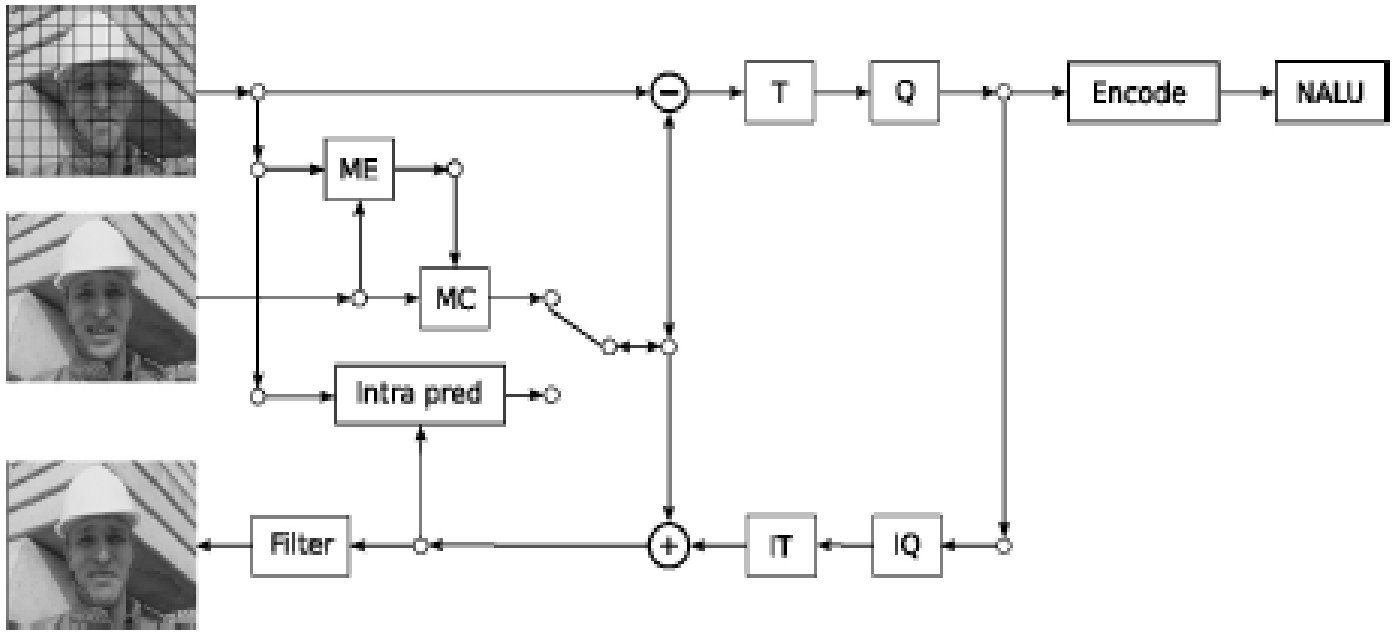


Fig. 4. H.264 compression overview.

- Therefore each scalable NALU belongs to a certain dependency layer (most commonly for resolution-scalability), a certain quality layer (to enable SNR scalability), and a certain temporal layer (to enable different frame rates).

III. Multimedia Encryption

- The classic application scenario of video encryption is in digital rights management (DRM), more precisely copyright protection, in which content providers aim to secure their business value, i.e., they want to prevent uncompensated redistribution of their content, very frequently videos.

- It is also common practice, that content providers, e.g., as frequently applied in pay-TV, offer free public access to parts of their (low quality) content to attract potential customers.

- In the application scenario of transparent encryption (referred to as perceptual encryption) the availability of a public low quality version is a requirement and the threat is that an attacker is able to compute a reconstruction of the original content with higher quality then the available public version [see Fig. 1(b)].

- Privacy preservation is also a concern in the context of video encryption, e.g., a commonly referred application is privacy preserving video surveillance; here the privacy of the people and objects in the video should be preserved; analogous problems for still images is currently facing Google with its StreetView application.

- The security threat in privacy preserving surveillance is the identification of a human person or object, e.g., a license plate, in the video, which thus has to be prevented (see Fig. 2).
- Video conferences are another prominent application scenario in which video data is encrypted.

- Though not a distinct application scenario, mobile computing is often referred to in the context of multimedia encryption, as the lower performance of mobile devices imposes strict constraints on the computational complexity, which is an argument for low-complexity encryption approaches.

A. Security/Quality/Intelligibility

- While conventional cryptographic security notions are built upon the notion of message privacy (referred to as MP-security), i.e., nothing of the plaintext message can be learnt/computed from the ciphertext message, the privacy of the message (video) is of limited concern for the content providers, but a redistribution of a sufficient quality version poses the threat (to their business model).

- The security of the video cryptosystem has to be defined with respect to this threat, i.e., the reconstruction of a sufficient quality version on the basis of the ciphertext, which leads to a specific security notion for multimedia encryption, which we refer to as MQ-security (message quality security) in this talk.

- A video is encrypted and an adversary must not be capable to compute a reconstruction of the plaintext with higher quality than allowed in the application scenario. (approximation attack!)
- It is sufficient for DRM scenarios that the quality is severely reduced, such that a redistribution is meaningless.

- In the context of privacy-preservation the quality/intelligibility of a video is measured in terms of recognizability of faces and persons.

- If one considers the raw video data as plaintexts the preservation of any information, even the preservation of the length of the compressed video stream or the length of units that comprise the compressed video stream constitutes a security violation, as even the compressibility of the raw video data leaks information on the raw video data.

- If encryption is conducted after compression the compressibility information is contained in the length of the compressed video data. This security notion appropriately models the security requirements of highly confidential video communications, e.g., video conferences in politics and economy.

- The preservation of format-compliance could be assumed to compromise security, recent contributions from the cryptographic community [5] discuss the topic in depth and define a concise formal framework and re-formulate the MP-security notion for format-preserving encryption (MP-security is defined for equal length format-compliant data) and also analyze format-preserving encryption algorithms, which are proved to be secure.

- [4] T. St¨utz and A. Uhl, “Efficient format-compliant encryption of regular languages: Block-based cycle-walking,” in *Proc. 11th Joint IFIP TC6 and TC11 Conf. CMS*, IFIP Advances in Information and Communication Technology, vol. 6109. May 2010, pp. 81–92.
- [5] M. Bellare, T. Ristenpart, P. Rogaway, and T. Stegers, “Format preserving encryption” in *Proc. SAC*, vol. 5867. Aug. 2009, pp. 295–312.

- 1) *Lightweight/Soft/Partial/Selective Encryption:* Some contributions to multimedia encryption propose the application of less secure but more efficient encryption algorithms (soft encryption), i.e., the computational complexity to break the employed cryptosystem with respect to MP-security is limited.

- E.g., in [6] it is proposed to employ a weaker cipher (an AES derivative with fewer rounds) for the less important parts of the bitstream. Often obviously insecure algorithms are employed (e.g., adding constants to the coefficients [7]) which also fall into that category.

- [6] Y. Fan, J. Wang, T. Ikenaga, Y. Tsunoo, and S. Goto, “A new video encryption scheme for H.264/AVC,” in *Proc. Adv. Multimedia Information Process.*, 2007, pp. 246–255.
- [7] H.-J. Lee and J. Nam, “Low complexity controllable scrambler/ descrambler for H.264/AVC in compressed domain,” in *Proc. ACM Multimedia*, 2006, pp. 93–96.

- Another approach to reduce the computational complexity of encryption is selective/partial encryption of the bitstream with a secure cipher [8].
- [8] A. Massoudi, F. Lef`ebvre, C. De Vleeschouwer, B. Macq, and J.-J. Quisquater, “Overview on selective encryption of image and video, challenges and perspectives,” *EURASIP J. Inform. Security*, vol. 2008, no. 179290, p. 18, 2008.

- In this talk, we will discuss the schemes in a cipher independent fashion, i.e., all encryption proposals will be considered to employ the same secure cipher as single source of pseudo-randomness.

B. Preserved Functionality

- Nonscalable and scalable H.264 bitstreams are accessible via a network abstraction layer (NAL), i.e., the coded video data is a sequence of separate NAL units (see Fig. 3 for a possible mapping of raw video data to NAL units).

- H.264 bitstreams are embedded into *Container* formats for transmission and storage, and depending on the encoding settings, bitstreams allow certain operations, such as extraction of IDR-picture (comparable to an I-frame in previous standards), extraction of a subset of the frames, cropping and in the case of SVC the extraction of substreams with different spatial resolution, temporal resolution and SNR quality.

- A wide range of watermarking algorithms specifically tailored to H.264 have been proposed, e.g., [9], [10]–[14] and a joint application of encryption and watermarking, especially watermarking encrypted content, is often desirable [11], [13], [14], [15].

- [9] R. Iqbal, S. Shirmohammadi, and A. El-Saddik, “Secured MPEG-21 digital item adaptation for H.264 video,” in *Proc. ICME*, Jul. 2006, pp. 2181–2184.
- [10] S. Lian, Z. Liu, Z. Ren, and H. Wang, “Commutative encryption and watermarking in video compression,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 6, pp. 774–778, Jun. 2007.
- [11] M. Noorkami and R. M. Mersereau, “Digital video watermarking in P-frames,” in *Proc. SPIE Conf. Security Steganography Watermarking Multimedia Contents IX*, vol. 6505. Jan. 2007.
- [12] D. Zou and J. A. Bloom, “H.264/AVC stream replacement technique for video watermarking,” in *Proc. IEEE ICASSP*, Mar. 2008, pp. 1749– 1752.

- [13] D. Zou and J. Bloom, “H.264 stream replacement watermarking with CABAC encoding,” in *Proc. IEEE ICME*, Jul. 2010, pp. 117–121.
- [14] P. Meerwald and A. Uhl, “Robust watermarking of H.264-encoded video: Extension to SVC,” in *Proc. 6th IIH-MSP*, Oct. 2010, pp. 82–85.
- [15] M. U. Celik, A. N. Lemma, S. Katzenbeisser, and M. van der Veen, “Lookup-table-based secure client-side embedding for spread-spectrum watermarks,” *IEEE Trans. Inform. Forensics Security*, vol. 3, no. 3, pp. 475–487, Sep. 2008.

- 1) *Format-Compliance:* A bitstream is denoted format compliant or H.264-compliant, if it suffices the syntax's and semantics' requirements of the H.264 standard. A format compliant H.264 bitstream has to be accepted by every H.264- compliant decoder without any undefined decoder behavior.

- The encrypted H.264 bitstreams where the encrypted data is signaled as supplementary data (e.g., using SEI messages) are still format-compliant, but the encrypted video data is treated completely different compared to plaintext video data.

- Thus, the application of conventional tools to process the video data may lead to unexpected and undesired behavior, e.g., rate adaptation algorithms may not perform optimal on the encrypted bitstreams. Thus we say that a functionality is preserved in a compliant fashion, if exactly the same processes as for an H.264 bitstream are applicable.

- 2) *Packetization: NAL Syntax/Structure:* The preservation of the NAL structure and syntax requirements enables the transparent application of *Standard Container* formats and tools for H.264.

- 3) *Fast Forward/Extraction of Subsequences/Scalability: NAL Semantics:* The additional preservation of the NAL semantics, i.e., the NAL unit type enables more sophisticated processing of the encrypted bitstream, such as fast forward, e.g., to the 100th IDR-picture of a coded video sequence, the extraction of subsequences, e.g., the last 10 min of a football match.

- In case of SVC this enables the preservation of scalability in the encrypted domain. The preservation of scalability in the encrypted domain allows us to even adapt the encrypted video. In SVC the scalability information is contained in the NALU headers, this information has to be preserved in the encrypted domain.

- In the context of DRM-systems the preservation of scalability allows to adapt the encrypted content to diverse hardware platforms efficiently without the need for the encryption key, e.g., a single content representation for mobile devices and home cinema systems.

- 4) *Transcodability:* Even if the bitstreams are not coded with SVC, the bitstreams can be transcoded via coefficient requantization and it is beneficial if this property is preserved in the encrypted domain [13].

- [13] N. Thomas, D. Lefol, D. Bull, and D. Redmil, "A novel secure H.264 transcoder using selective encryption," in *Proc. IEEE ICIP*, Sep. 2007, pp. IV-85–IV-88.

- 5) *Robust Watermarking:* Many watermarking algorithms are compression-format independent and require the raw video data as starting point. In the scope of this lecture we only consider proposals that have been explicitly designed for the joint application with H.264.

- Furthermore we do not discuss the watermarking algorithms in detail, but limit the discussion to the marking space and whether embedding could be conducted in the encrypted domain. The embedder may be required to have precomputed metadata for embedding.

- We consider two basic approaches for watermarking compressed H.264 bitstreams, watermarking of DCT coefficients [14], [15] and bitstream substitution watermarking which is conducted by via special modifications to the intraprediction modes for CAVLC [16] or by special modifications of MVDs for CABAC [17].

- [14] S. Lian, Z. Liu, Z. Ren, and H. Wang, “Commutative encryption and watermarking in video compression,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 6, pp. 774–778, Jun. 2007.
- [15] M. Noorkami and R. M. Mersereau, “Digital video watermarking in P-frames,” in *Proc. SPIE Conf. Security Steganography Watermarking Multimedia Contents IX*, vol. 6505. Jan. 2007.
- [16] D. Zou and J. A. Bloom, “H.264/AVC stream replacement technique for video watermarking,” in *Proc. IEEE ICASSP*, Mar. 2008, pp. 1749– 1752.
- [17] D. Zou and J. Bloom, “H.264 stream replacement watermarking with CABAC encoding,” in *Proc. IEEE ICME*, Jul. 2010, pp. 117–121.

- DCT coefficient watermarking requires a partial decoding, i.e., entropy decoding of the coefficients and afterwards the watermark is embedded via modifications of the DCT coefficients. The approach of [15] requires access to the entire coefficients, while [14] proposed a quantization-based watermarking scheme, which preserves the signs of the coefficients and thus can be combined with coefficient sign encryption.

- The interprediction mode watermarking approach [16] and the MVDs watermarking approach [17] can be applied by simple substitutions of bits in the compressed bitstream. Thus two basic watermarking approaches for H.264 are considered in this lecture:
- 1) DCT coefficient watermarking;
- 2) stream substitution watermarking.

- ROI encryption is useful for privacy preservation, i.e., the image areas in which persons and privacy sensitive objects appear have to be rendered unintelligible (encrypted).
- Privacy preserving encryption has gained significant interest [18], [19], [20] and our discussion in this lecture is deliberately only focused on the H.264 encryption details

- [18] F. Dufaux and T. Ebrahimi, “A framework for the validation of privacy protection solutions in video surveillance,” in *Proc. IEEE ICME*, Jul. 2010, pp. 66–71.
- [19] A. Senior, Ed., *Protecting Privacy in Video Surveillance*. Berlin, Germany: Springer, 2009.
- [20] F. Dufaux and T. Ebrahimi, “H.264/AVC video scrambling for privacy protection,” in *Proc. IEEE ICIP*, Oct. 2008, pp. 1688–1691.

- In conclusion, we distinguish the following distinct security and application scenarios.
- 1) *Highest level security (MP security on the raw video data):* Source independent length packets and MP-secure encryption schemes, i.e., practically AES encryption in a secure mode, can meet the imposed requirements.

- 2) *Content confidentiality/visual semantic security (MQ security with a security metric that can identify a security breach for the visual data):* Encryption of source dependent length packets can meet these requirements. Content confidentiality lowers the security requirements, in order to allow functionality, such as optimal adaptation, to be preserved (which leads to a security breach for highest level security).

- 3) *Sufficient encryption (MQ security with a quality metric that can reliably determine the subjectively perceived quality):* Many encryption schemes may offer this property, but the computational hardness of the problem (to recover a higher quality then targeted) has not yet been proved for any proposal.

- However, for most encryption schemes no practical attacks exist. Sufficient encryption even more than content confidentiality lowers the security requirements and even more functionality can possibly be preserved, such as transcodability and watermarking.

- 4) *Transparent/perceptual encryption (MQ security with a quality metric):* Similar to sufficient encryption, the goal of transparent encryption is to reduce the quality, but transparent encryption also requires that a certain quality is preserved, i.e., that the ciphertext can be decoded with a certain quality in order to provide a public low quality version.

- 5) *ROI encryption/privacy preserving encryption (MQ security with an intelligibility metric):* The security of a privacy preserving encryption scheme can be defined on the basis of intelligibility, i.e., does (automatic) face recognition work on the encrypted video [18].

D. Evaluation Criteria

- Given a video encryption scheme suitable for a security and application scenario, security has to be assessed in terms of the computational complexity to break the scheme with respect to the appropriate security notion. A video encryption scheme can have a negative impact on the compression performance, which is a major evaluation criterion for the practical applicability of an approach.

- 1) *Online/Offline Scenario:* The computational complexity of an encryption scheme also is a decisive factor for its applicability. Depending on the application context, the video data is available in a raw, uncompressed format and H.264 compression has to be conducted anyway (this is referred to as online scenario, e.g., video conferencing) or the video data is already available as compressed H.264 bitstream (this is referred to as offline scenario).

- The application context may further require that certain properties and associated functionalities of the video bitstream are preserved in the encrypted domain. Thus the main assessable properties of a video encryption scheme in a security and application scenario are as follows:

- 1) security (in a certain application scenario with respect to the specific security notion);
- 2) compression efficiency;
- 3) computational complexity;
- 4) preserved functionality (format-compliance, packetization, scalability, transcodability, and watermarking).