Scalable encryption for multimedia content access control

Hong Heather Yu
Panasonic Information and Networking Technologies Laboratory
Heathery@research.panasonic.com

Abstract

Traditional cryptography systems treat every portion of a message (a video, an image, or an audio) equally and encrypt the entire message as a whole. As a result, those security systems often have only two states: access authorization and access denial. To facilitate multi-level access control with interoperability, scalable security mechanism is needed. Further, the availability of varying network bandwidth and diverse receiver device capabilities demand scalable and flexible approaches that are capable of adapting to changing network conditions as well as device capabilities. In this paper, we describe a multimedia encryption scheme that supports scalability. One desirable feature of the proposed scheme is its simplicity and flexibility in supporting scalable content access control.

1 INTRODUCTION

In the digital world, content protection is an important topic. There is increasing application need for digital content access control technologies. One application interested by many content distributors is to support multiple levels of access capability with flexible scales and to allow different users with different level of rights to have different level of access capability. For example, general customers may preview a video in is very coarse scale, a club member may preview it in a finer scale, and one may view it in its finest quality after paying the access fee (see Figure 1.) Moreover, it is often important to support interoperability. For instance, different content distributors may give different scale of accessibility to different level of customers while the same playback device is used by the same customer to access contents from different distributors. Traditional encryption schemes do not provide such capabilities. What kind of improvement maybe needed to attain scalability as well as flexibility for scalable multimedia content access control? In the following section of this paper, we propose a scalable encryption scheme for secure multimedia distribution and scalable content access control. Discussion and future work direction will be presented in the last section.

2 SCALABLE ENCRYPTION

2.1 Existing techniques

In a traditional communication system, the encoder compresses the source media into a fixed bit rate that may be equal to or less than the channel capacity and sends it to the receiver. Given the assumption that the receiver can obtain and decode all the bits in time, the receiver reconstructs the media using all the bits received. Similarly, in a traditional encryption system, the sender encrypts the entire message (all the bits of the entire video or audio) and sends it to the receiver. After receiving the entire encrypted message successfully, the receiver decrypts it and plays it. In such a system, two or more assumptions are made: the channel capacity is known; the receiver can receive all the bits in time for decryption; and the receiver is capable of reconstructing the media fast enough. This kind of traditional security system often has only two states of accessibility: access authorization and access denial. To facilitate multi-level access control and multi-distributor interoperability, scalable security mechanism is needed. Furthermore, those assumptions are challenging in many multimedia applications due to the unknown and often varying channel capacities, unknown and diverse processing power and storage capacities of receiver devices, and often high computational complexity of conventional encryption algorithms. To provide scalable access of multimedia, scalable and fine granularity scalable (FGS) [4][5][6] video compression algorithms have been proposed such that a bitstream is partially decodable at any bit rate within a bit rate range to reconstruct the medium signal with the optimized quality at that bit rate.

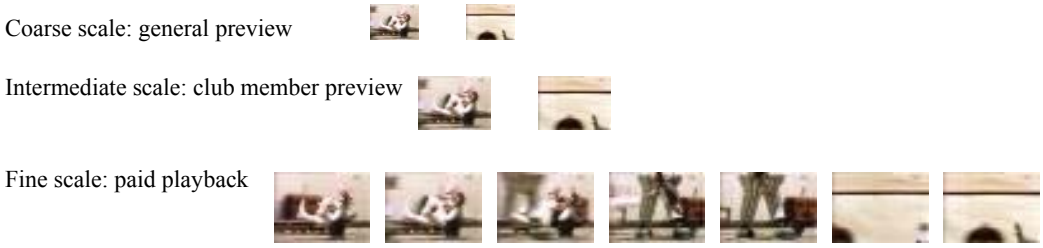

FIGURE 1. Sample application of scalable content access control

However, if a medium compressed using scalable coding or FGS coding needs to be protected and previous non-scalable cryptography algorithms are used (the bitstream is encrypted uniformly), the advantages of scalable coding and FGS coding may be lost. In the mean time, to resolve the real time constraint in many applications, selective encryption was proposed [1][2][3]. That is only some parts of the entire bitstream are encrypted while the rest are left in the clear. For instance, only I-frames or I-frames plus the I-blocks in P and B frames of a MPEG video are encrypted. While selective encryption provides one additional preview access level, the design focus of the proposed algorithms is real time decoding instead of compatibility with scalable coding. The non-scalability nature of the cryptography algorithms makes it hard to reconstruct the medium at any bit rate for transmission through various networks and playback on various kinds of devices with flexible access specification.

2.2 Scalable encryption (SE)

Figure 2 illustrates SE scalability. The enhancement layer is progressively, e.g., bit-plane by bit-plane, decodable (decryptable) such that it is able to perform real time tradeoffs between the various SNR-temporal-spatial-complexity scalabilities at transmission, processing, or post-processing. The principal idea of SE is to simulate FGS coding structure of the data while encrypting the data sequence with suitable security level. Figure 3 shows the general system architecture. The bitstream of the enhancement layer may be truncated into any number of bits and the decoder should be able to reconstruct the bitstream from the base layer and the truncated enhancement layer with the quality/resolution of the image proportional to the number of bits decoded, i.e., having FGS scalability similar to that was illustrated in Figure 3. Due to page limitation, we will focus only on JMOVIE formatted or noncompressed video (i.e., no motion vector) in this paper. The basic algorithm will be presented here.

For a JMOVIE formatted video, each frame of the video I is first partitioned into base layer A and enhancement layer B in the DCT frequency domain. The base layer consists of the DC coefficient or the DC plus low band AC coefficients. The rest AC coefficients are used to construct the enhancement layer. The base layer and the enhancement layer are in general encrypted separately since it is assumed the lowest resolution accepted at any receiver is the one reconstructed from the entire base layer bitstream A. The base layer can be encrypted using any suitable encryption algorithms, for instance, sign encryption and block scrambling for low computational overhead and DES for moderate computational overhead. The enhancement layer that provides FGS capability is encrypted as following to preserve the scalability.

Assume N bands of 8X8 blocks with diagonal zigzag scan from the lowest band to the highest band and from the upper left block to the lower right block of the absolute values as following:

Band 1:
9 3 4 2 5 0 0 0 1 2 … 0 0

Band 2:
10 0 6 0 0 3 0 2 2 0 … 0 0
… …

Band N:
12 8 5 4 2 0 3 0 2 0 … 0 0

Writing them in binary forms 4 bit-planes:
1 0 0 0 0 0 0 0 0 0 … 0 0 ---- (B1 MSB <- B(1,3))
0 0 1 0 1 0 0 0 0 0 … 0 0 ---- (B1 MSB-1 <- B(1,2))
0 1 0 1 0 0 0 0 0 1 … 0 0 ---- (B1 MSB-2 <- B(1,1))
1 1 0 0 1 0 0 0 1 0 … 0 0 ---- (B1 MSB-3 <- B(1,0))

1 0 0 0 0 0 0 0 0 0 … 0 0 ---- (B2 MSB <- B(2,3))
0 0 1 0 0 0 0 0 0 0 … 0 0 ---- (B2 MSB-1 <- B(2,2))
1 0 1 0 0 1 0 1 1 0 … 0 0 ---- (B2 MSB-2 <- B(2,1))
0 0 0 0 0 1 0 0 0 0 … 0 0 ---- (B2 MSB-3 <- B(2,0))

… …
1 1 0 0 0 0 0 0 0 0 … 0 0 ---- (BN MSB <- B(3,3))
1 0 1 1 0 0 0 0 0 0 … 0 0 ---- (BN MSB-1 <- B(3,2))
0 0 0 0 1 0 1 0 1 0 … 0 0 ---- (BN MSB-2 <- B(3,1))
0 0 1 0 0 0 1 0 0 0 … 0 0 ---- (BN MSB-3 <- B(3,0))

Converting them into (RUN, EOP) symbols, we have:

(0,1) - (B1 MSB)
(2,0),(1,1) - (B1 MSB-1)
(1,0),(1,0),(5,1) - (B1 MSB-2)
(0,0),(0,0),(2,0)(3,1) - (B1 MSB-3)

(0,1) - (B2 MSB)
(2,1) - (B2 MSB-1)
(0,0),(1,0),(2,0),(1,0),(0,1) - (B2 MSB-2)
(5,1) – (B2 MSB-3)
… …

(0,0),(0,1) – (BN MSB)
(0,0),(1,0),(0,1) – (BN MSB-1)
(4,0),(1,0),(1,1) – (BN MSB-2)
(2,0),(3,1) – (BN MSB-3)

The M (RUN, EOP) symbols $B(x, y, j)$ = (0,1);(2,0),(1,1);(1,0),(1,0),(5,1);(0,0),(0,0),(2,0)(3,1);(0,1);(2,1);(0,0),(1,0),(2,0),(1,0),(0,1);(5,1);… …,(0,0),(0,1);(0,0),(1,0),(0,1);(4,0),(1,0),(1,1);(2,0),(3,1); are then scrambled on bit plane level using key K of length 128bit or longer to generate a scrambled sequence

$$B' = Enc_K(B) = \underset{j}{Enc_K}(B(x, y, j)) \quad = \quad (0,0),(0,1),(2,0),$$

(1,1),(1,0),(1,0),(5,1),(5,1),(0,1),(0,0),(1,0),(0,1),(2,0),(1,0),(1,0),(0,0),(0,1),(0,0),(0,0),(2,0)(3,1),… …,(0,1),(2,1),(4,0),(1,0),(1,1),(2,0),(3,1) with $K=K1+K2$, $x=[1,N]$, and $y=[0,3]$. Because sign bit scrambling will not introduce bit rate increase, an additional layer or sign

bit scrambling can be used to increase the security level. The enhancement layer bitstream is transmitted in an MSB-first LSB-last fashion to the receivers.

In some applications, the enhancement layer bitstream input of the decoder *B"* is a truncated version from that of the output of the encoder *B'*. The length of each bitstream received by different devices may vary a great deal. To decode the truncated bitstream *B",* EOP symbols are first found. If the last (RUN, EOP) symbol received is an (x,0) instead of an (x,1), the decoder looks back L symbols until it finds an (x,1) symbol and discard the last L-1 symbols. Decoder then unscramble the curtailed bitstream with decoding key K'

$$B^* = Dec_{K'}(B'') = \underset{j}{Dec_{K1'}}(B''(x', y', j'))$$

and decompress the bitstream. Since the base layer are not to be truncated, it is decrypted in the frequency domain after dequantization and the base layer image is then merged with the enhancement layer image for playback. Because the enhancement layer bitstream are received MSB-first and LSB-last with bit-plane level encryption, the descrambled shortened bitstream can be used to enhance resolution/quality of the base layer image proportional to the number of bits received.

To achieve fine grained scalability on non-compressed video, two kinds of schemes can be used. To control computational overhead to the minimum amount, bit-plane scrambling can be used. Assume the *s*th bit value of the *t*th pixel of the *i*th frame of an M-bit video in an zigzag scan is $B_i(t,s)$, $b_i' = \underset{t}{Enc_K}(b_i(t,s))$ for s=[0,R] = [MSB,MSB-R] can be performed. For partial encryption, R<M-1 where full encryption requires R=M-1. In the event an truncated bitstream of L=*rN*+q bits of an *N=N1xN2* video frame is received, *m* and *q* are both integers with *q*<*m*, the *rN* bits are used to reconstruct the video frame with corresponding resolution at receiver. Notice that to have fine grained

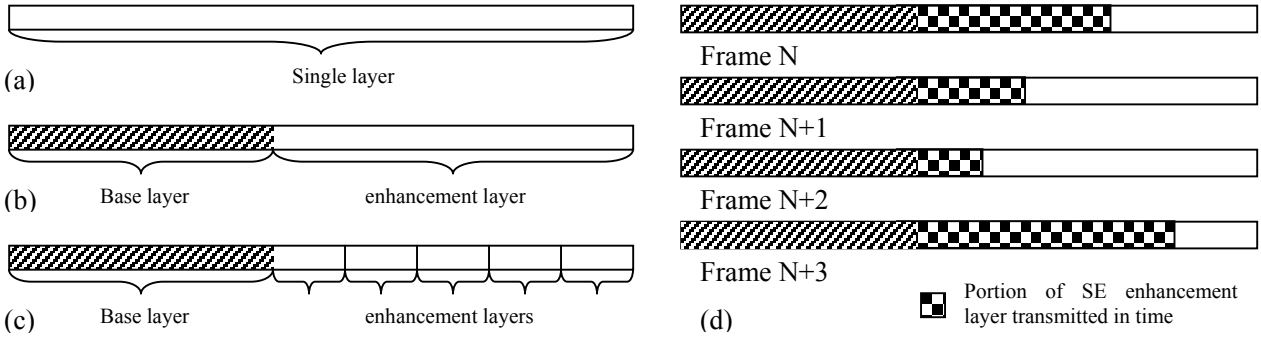

FIGURE 2. Illustration of sample FGS scalability compared with non-scalable and two-stair scalable coding. (a) Non-scalable; (b) scalable; (c) FGS scalable; (d) FGS scalability in a multiple access level application.

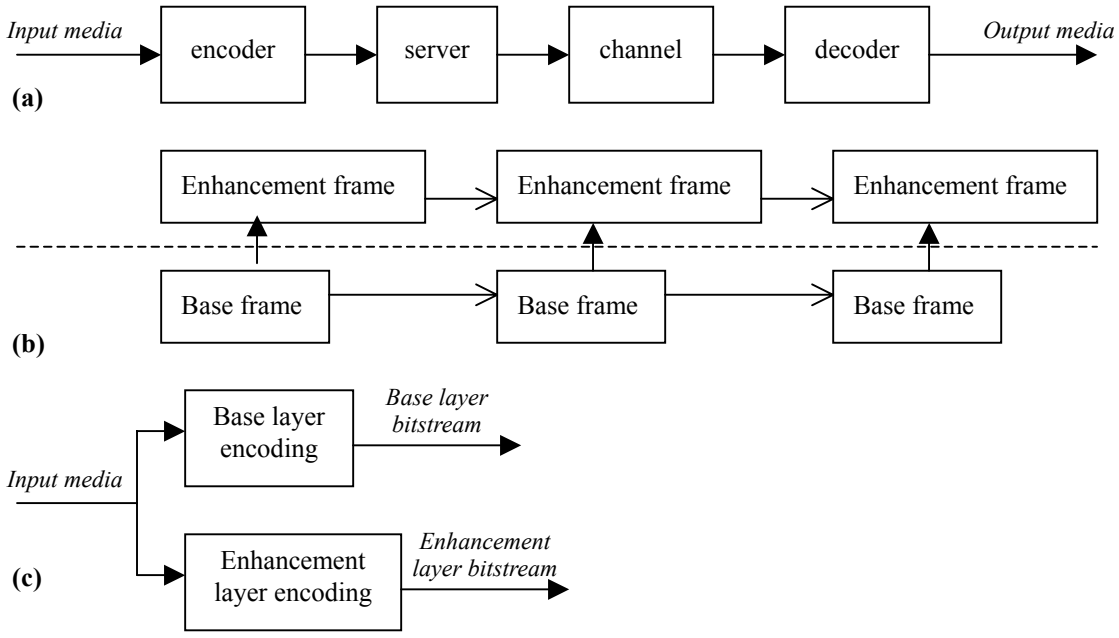

FIGURE 3. (a) general system structure; (b) temporal scalable; (c) spatial scalable encoder;

scalability, the encrypted bitstream has to be transmitted in a way that the MSBs of all frames shall be received by all receivers, then MSB-1s and so on and so forth. While this algorithm is plausible for its simplicity, a second method that utilizes Haar Wavelet transformation may give better bit number – quality - scalability tradeoff with the penalty of additional computation power. First, Haar Wavelet decomposition is performed on each input video frame *I*. The lowest band L^{z+2} constructs the base layer bitstream *A* while the rest forms the supplement layer bitstream $\mathcal{S}$. Similar to that was discussed earlier, *A* maybe encrypted using any suitable encryption algorithms. The supplement layer maybe encrypted using multilayered encryption to achieve fine grained scalability: for z=[1,Z] level decomposition, $\mathcal{S}^z = Enc_{Kz}(L^zHL,\ L^zLH,\ L^zHH)$, $\mathcal{S} = \mathcal{S}^1, \mathcal{S}^2, \ldots \mathcal{S}^Z$, where *Enc* can be intra-band coefficient scrambling, intra-level coefficient scrambling, intra band bit-plane scrambling, or any application suitable encryption algorithms. Alternatively, a bit-plane by bit-plane scrambling algorithm similar to that was discussed in JMOVIE enhancement layer encryption can be used.

2.3 Scalable content access control

In the case of scalable content access control, different receiver with different access authorization will receive different keys via the secure transmission channel. In accordance, the enhancement layer is partitioned into several sub-layers with the number of layers equals to the number of access levels minus 1. With an elegant key management system, multi-level access control and multi-distributor interoperability can be achieved on flexible scales.

3 RESULT, DISCUSSION AND FUTURE WORK

The proposed SE algorithms provides high scalability with lower computational overhead compared with the conventional encryption algorithms. Because when a encryption/decryption key is <128bit, it is considered computationally feasible for an attacker to decode the message using brute force attack, the key length has to be 128 or longer. 50%, 75%, and 100% bit-plane level scrambling are tested on video. The computation overhead is less than 20% of that full encryption using pre- or post-quantization coefficients scrambling in all cases. For JMOVIE formatted video, the bitrate overhead is less than 10% on average. We noticed that when a small percentage of the bit-planes are scrambled, some level of perception of the video content is possible, e.g., a person appeared in the first part of the video and one more person joined later in the video. In high security required applications, the number of scrambled bit-planes should be as large as possible. However, because the level of perception quality is so low, in many commercial applications such as DTV and Digital Cinema, it will not cause any lost of revenue. It would be feasible to use to trade for lower cost.

The most significant difference between SE and conventional encryption algorithms is that SE considers each coefficient as a binary number of several bits instead of a decimal integer of a certain value and scrambling is done on bit plane level to provide fine granularity scalability.

(a) original (b) <50% scrambled (c) 70% scrambled

FIGURE 4. A sample video frame

Encryption on the entire bitstream using conventional secure encryption algorithms such as DES, triple DES, and AES will no doubt provide the highest security levels. However, due to the need for large amount of processing power, they are not applicable in some time-constrained applications. Considering the value of commercial video and audio decreases significantly after a period of time, light weighted partial encryption algorithm with scalability and low computational complexity can provide adequate security level with minimum increase on bit rate and processing time. How to analyze the rate-complexity-security tradeoffs and find the optimum algorithm for each given rate-complexity, complexity-security, and rate-security pairs is an important future work. Furthermore, SE algorithms that are compliant to wavelet zero-tree based fine granular scalable coding and matching pursuit coding of image residue are being investigated.

REFERENCE:

[1] I. Agi, L. Gong, "An empirical study of MPEG video transmissions", in Proc., Internet Society Symp. On Network and Distributed Systm Security, Feb., 1996

[2] Y. Li, Z. Chen, S. Tan, R. Campbell, "Security enhanced MPEG player", in Proc., IEEE First Int. Workshop on Multimedia Software Development, Mar., 1996

[3] L. Qiao, K. Nahrstedt, "A new algorithm for MPEG video encryption", in Proc., the First Int. Conf. On Imaging Science, Systems and Tech., July 1997

[4] R. Aravind, M. R. Civanlar, A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithm", *IEEE Trans. on Circuits Syst. Video Tech.*, vol 6, Oct, 1996

[5] H. Gharavi, M. H. Partovi, "Multilevel video coding and distribution architectures for emerging broadband digital networks", *IEEE Trans. on Circuits Syst. Video Tech.*, vol 6, Oct, 1996

[6] W. Li, Over view of fine granularity scalability in MPEG-4 video standard, *IEEE Trans on Circuits and Syst. Video Tech*. V 11, N 3, 2001