SECURITY EVALUATION FOR COMMUNICATION-FRIENDLY ENCRYPTION OF MULTIMEDIA

Yinian Mao and Min Wu

ECE Department, University of Maryland, College Park

ABSTRACT

This paper addresses the access control issues unique to multimedia, by using a joint signal processing and cryptographic approach to multimedia encryption. Based on three atomic encryption operations, we present a systematic study on how to strategically integrate different atomic operations to build a video encryption system. We also propose a set of multimedia-specific security metrics to quantify the security against approximation attacks and to complement the existing notion of generic data security. The resulting system can provide superior performance to both generic encryption and its simple adaptation to video in terms of a joint consideration of security, bitrate overhead, and communication friendliness.

1. INTRODUCTION

The burgeoning development in digital multimedia and communication technologies has paved ways for people around the world to acquire, utilize, and share multimedia content. For the wide availability of multimedia information and successful commercialization of many related services, assuring that the multimedia information is used only by authorized users for authorized purposes has become essential. This paper discusses about protecting the confidentiality and achieving access control for multimedia information, with the emphasis on system integration and security evaluation.

Content confidentiality and access control is generally addressed by encryption. In principle, digital multimedia can be encoded into a bitstream and encrypted in the same way as generic data [1]. By doing so, however, the encryption will wipe out the inherent structures and the syntax of multimedia data. Many functionalities provided by state-of-the-art signal processing for multimedia will disappear after encryption, such as scalable coding, unequal error protection, and compressed domain search and indexing [2, 3]. Consequently, a number of schemes have been proposed to take signal processing into consideration when encrypting multimedia [3, 4]. Among these schemes, transform/codeword domain shuffling and codeword domain index encryption are some common choices.

Due to the unique nature of multimedia data, the "all-or-nothing" protection in generic data security is not always appropriate for measuring the security of multimedia encryption [3, 4]. Beyond the exact recovery from ciphertext, it is important to ensure partial information that is perceptually intelligible is not leaked out from the ciphertext. The perceptual aspects of the security should also take into account approximation attacks from adversaries who employ prior knowledge and correlation in multimedia data to produce an estimated version of the plaintext.

Given the prevalence of multimedia content and the issues related to multimedia encryption as mentioned above, there is need to study the encryption of multimedia in a systematic way. Our goal is to design encryption systems for multimedia that are friendly to communication and signal processing techniques, reduce the cost of such systems, and achieve an appropriate level of security. In [5], we proposed two atomic encryption operations for multimedia and provided analytical results on the bitrate overhead of the encrypted data. In this paper, we first propose a notion of multimedia-specific security and two quantitative security metrics. We then show how to integrate different encryption operations to build a video encryption system and discuss the tradeoff among security, compressibility, and compatibility to intermediate processing during transmission.

The rest of the paper is organized as follows. After introducing the proposed security metrics in Section 2, we provide brief reviews of the encryption operations in Section 3. Section 4 discusses video encryption system design and shows the results under different settings. Finally, conclusions are drawn in Section 5.

2. MEASURING THE SECURITY FOR MULTIMEDIA ENCRYPTION

In the literature, there has been extensive discussion on the security of generic data encryption. The investigations in the security issues specific to multimedia, however, remains limited. In this section, we discuss why the security of multimedia encryption needs extra attention, and how can we evaluate the security beyond the bit-by-bit representation of media data. We also propose two visual security metrics that incorporate the perceptual apsects in visual data.

2.1. Exact Knowledge Versus Approximation

In generic data encryption, the notion of practically provable security was introduced in [6] to measure the security against attackers' recovery of the exact plaintext from the ciphertext. This notion quantifies the security strength of a system in terms of the amount of resources needed to break the system. Assume that an adversary has the constraint t on the computing time, and q on the number of queries he/she can make to an encryption oracle. The notion of $(t, q; \epsilon)$-security indicates that the success probability for this adversary is at most ϵ when his/her resources are bounded by the constraints mentioned above.

Due to the spatial and temporal correlation of multimedia, the encrypted content may be approximately recovered based on the syntax, context, and the statistical information known as *a priori*. This is possible even when the encrypted part is provably secure according to the generic security notion. For example, in MPEG video encryption, when motion vector fields are encrypted and cannot be accurately recovered, a default value 0 can be assigned to all motion vector fields [3]. This approach results in a fairly good approximation for slow-motion frames. Additionally, the statistical information, neighborhood patterns, and smoothness

Authors' email: {ymao, minwu}@eng.umd.edu

criterion can help estimate an unknown area in an image and automatically reorder shuffled image blocks [8]. It is therefore important to introduce a notion of multimedia-specific security. Under such a notion, the possible information leakage should be evaluated against the approximation recovery in addition to that from the encrypted data.

2.2. Visual Security Metrics

Studies on human visual system suggest that two important types of information are extracted by an observer of a given image [7]. The first type is the edge and contour information, which describes the shape of the objects. The second type is the luminance or color space information. Based on this observation, we introduce an edge similarity score and a color similarity score to quantitatively measure the distance between two images. For gray scale images, their color similarity becomes luminance similarity, which will be detailed below.

Edge Similarity Score (ESS) The edge similarity score measures the degree of resemblance of the edge and contour information between two images. To evaluate edge similarity, two images of the same size are first divided into blocks. If the two images are not of the same size, they are resized and aligned by preprocessing modules. Then edge detection is performed for each block. The dominant edge direction is extracted and quantized into one of the eight representative directions. The representative edge directions have equal angular distance of 22.5 degrees between two neighbor directions in a polar coordinate system. We use indices 1 to 8 to represent these eight directions, and use index 0 to represent a non-edge block. Denoting e_{1i} and e_{2i} as the edge direction indices for the i-th block in two images, respectively, the edge similarity score (ESS) for a total of N image blocks is computed as:

$$ESS = \frac{\sum_{i=1}^{N} w(e_{1i}, e_{2i})}{\sum_{i=1}^{N} c(e_{1i}, e_{2i})}. \quad (1)$$

Here, $w(e_1, e_2)$ is a weighting function defined as

$$w(e_1, e_2) = \begin{cases} 0 & \text{if } e_1 = 0 \text{ or } e_2 = 0, \\ |cos(\phi(e_1) - \phi(e_2))| & \text{otherwise,} \end{cases}$$

where $\phi(e)$ is the representative edge angle for an index e, and $c(e_1, e_2)$ an indicator function defined as

$$c(e_1, e_2) = \begin{cases} 0 & \text{if } e_1 = e_2 = 0; \\ 1 & \text{otherwise.} \end{cases}$$

The score ranges from 0 to 1, where 0 indicates that the edge information of the two images is highly distinct and 1 indicates a match between the edges in the two images. A special case arises when both images in comparison are very smooth, leading the denominator in (1) to 0. Although this is a match case, we assign an ESS score of 0.5 to it, because there is not much edge information extracted from either image. In our experiments, we partition the input images into non-overlapping 8x8 blocks and use the Sobel operator for edge detection.

Luminance Similarity Score (LSS) To capture the coarse luminance information, we introduce a block-based luminance similarity score. After the original images are partitioned into blocks, the average luminance values of the i-th block, y_{1i} and y_{2i}, from both images are calculated. We define the luminance similarity score as

$$LSS = \frac{1}{N} \sum_{i=1}^{N} f(y_{1i}, y_{2i}). \quad (2)$$

Here, the function $f(y_1, y_2)$ for each pair of average luminance values is defined as

$$f(y_1, y_2) = \begin{cases} 1 & \text{if } |y_1 - y_2| < \frac{\beta}{2}, \\ -\alpha \text{ round}(\frac{|y_1 - y_2|}{\beta}) & \text{otherwise,} \end{cases}$$

where the parameters α and β control the sensitivity of the score. Along with block-based aggregation, the α factor within the range from 0 to 1 and the quantization parameter β provide resistance to minor perturbation and noise. In our experiments, α and β are set to 0.1 and 3, respectively. A negative LSS value indicates substantial dissimilarity in luminance between two images.

3. OVERVIEW OF ENCRYPTION OPERATIONS

Atomic encryption operations are basic building blocks for encrypting multimeida. All the operations we review here satisfy the property of syntax preservation, namely, after applying these encryptions, the encrypted media still preserves the syntax prescribed by multimedia coding standards. As a result, many of the communication and signal processing techniques designed for unencrypted multimedia can also be applied to the encrypted data.

We first introduce the idea of *Generalized Index Mapping* [5, 3], which can be applied directly to symbols that take values from a finite set. Examples may include working with quantized coefficients, quantized prediction residues, and run-length coding symbols. The encryption process produces a ciphertext symbol $X^{(enc)}$ from a plaintext symbol X: $X^{(enc)} = T^{-1}[Encrypt(T(X))]$, where $T(\cdot)$ represents a codebook that establishes a bijective mapping between symbol values and indices represented by binary strings. The decryption process has a similar structure: $X = T^{-1}[Decrypt(T(X^{(enc)}))]$. In [5] and [9], we have developed analytical results on the bitrate overhead brought by index encryption, and have shown that by partitioning the input symbol range S into multiple subsets and restricting the encryption output to be in the same subset as the input symbol, the bit-rate overhead can be reduced at the expense of a reduced complexity for brute force attack.

Fine granularity scalability (FGS) is desired in multimedia communications to provide a near-continuous tradeoff between bitrate and quality. In [5], we proposed an *Intra Bitplane Shuffling* operation that is compatible with bit-plane coding, such as the recently adopted MPEG-4 FGS. To encyrypt the FGS layer video, random shuffling is applied in the transform coefficient domain on each bitplane of n bits. The shuffled bitplane will then be encoded using run-length coding. Using such an encryption approach, the scalable coded video can be protected without the loss of scalability in the encrypted bitstream, while maintaining a low bitrate overhead.

A multimeida coding system often partitions an input signal into segments and encodes each segment into a self-contained unit. Shuffling the order of such units according to a cryptographically strong permutation table has the advantage of preserving the compressibility as well as the syntax of the coded bitstream [1, 3]. We shall refer to such operations as *block shuffling*. A major drawback for block shuffling is that an attacker can exploit the correlation across the blocks, such as the continuity of edges and similarity of colors and textures, and reassemble the shuffled blocks with a much smaller effort than that of a brute force search [8]. Therefore, block shuffling alone is often not a secure encryption operation. However, as a complementary building block, it can help achieve good visual/auditory scrambling effect for multimedia data.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we present the design and evaluation of a video encryption system. We show the results of different system settings using the encryption operations reviewed in the last section. Four video clips are used in our experiment, namely, *Football*, *Coast-guard*, *Foreman*, and *Grandma*. Each video clip is 40 frames long and coded with MPEG-4 standard. The GOP size is set to 15 and all predictive frames are P frames. For these video clips, we identify three possible components in the base-layer video to which we apply the generalized index mapping. These components are: (1) the DC prediction residues of intra-blocks, (2) the motion vector (MV) residues of inter-blocks, and (3) a part of non-zero AC coefficients of intra-blocks. In addition, we also incorporate the random shuffling of macroblocks from both intra and predictively coded pictures.

There are 6 encryption setting and 3 approximation attack settings in our experiments. Encryption settings E1-E3 are listed below, where the encryption of DC, AC, and/or MV is based on the proposed generalized index mapping. The DC and AC encryption ranges are chosen as [-63,64] and [-32,32] with set partitioning, respectively [5]. Settings E4-E6 correspond to E1-E3 plus macroblock shuffling in the compressed bit-stream, respectively.
(E1) encrypting intra block DC residue by index mapping;
(E2) encrypting inter block MV residue in the first two PVOPs immediate following an IVOP, and all intra block DC residue;
(E3) encrypting all the components listed in E2, plus the first two non-zero AC coefficient of intra block.

Corresponding to the encryption settings, the settings for approximation attacks (A1-A3) that emulate an adversary's action:
(A1) set all intra block DC coefficients to 0;
(A2) set all intra block DC coefficients to 0 and set the encrypted motion vector values to 0;
(A3) including all the approximations in A2, plus set the encrypted AC coefficients to 0.

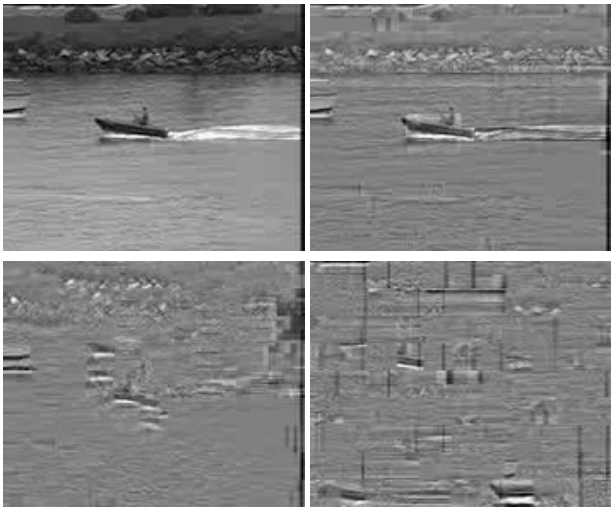

Fig. 1. Encryption results for *Coast-guard*. The encryption-approximation settings are: (top row, left to right) unencrypted, E1+A1; (bottom row, left to right) E2+A2, E4+A1.

4.1. Perceptual Security Against Approximation Recovery

Using the notion of multimedia-specific security for image and video and the proposed perceptual similarity scores in Section 2, we evaluate the perceptual security of different encryption configurations against approximation recovery attacks. We denote the application-dependent thresholds for ESS and LSS as ESS_{th} and LSS_{th}, respectively. An encrypted image/video is said to *pass* an edge or luminance similarity test against a certain attack if the resulting image/video from the attack has edge and luminance similarity scores lower than ESS_{th} and LSS_{th}, respectively. In the following experiment, we set ESS_{th} to 0.5 and LSS_{th} to 0, which we have found to provide sufficient security for video in many applications.

Table 1. Perception based security measures for video encryption

	Football		*Grandma*		*Foreman*	
Settings	ESS	LSS	ESS	LSS	ESS	LSS
E1+A1	0.70	-0.78	0.64	-2.13	0.71	-1.42
E2+A2	0.53	-0.85	0.46	-2.13	0.43	-1.48
E3+A3	0.53	-0.86	0.30	-2.13	0.40	-1.48
E4+A1	0.12	-0.93	0.05	-2.13	0.07	-1.47
E5+A2	0.13	-0.92	0.05	-2.13	0.06	-1.45
E6+A3	0.12	-0.92	0.04	-2.13	0.05	-1.47

Table 2. Relative Compression Overhead of the Encrypted Videos

	Football	*Foreman*	*Coastguard*	*Grandma*
E1	1.29%	1.75%	3.15%	6.96%
E2	3.88%	6.41%	8.74%	11.11%
E3	6.47%	9.62%	11.54%	24.61%

Table 1 lists the average ESS and LSS scores of three videos after approximation recovery. From the average LSS scores, we can see that the luminance information is well protected after DCs are encrypted and the score remains at a similarly low level as more video components are encrypted. However, from the average ESS scores we can also see that edge and contour information needs more protection than luminance information and block shuffling is an effective tool.

To examine the detailed ESS scores, we plot the frame-by-frame ESS score of *Coast-guard* under different encryption-attack settings in Fig. 2. The top curve is from the attacked video with DC encrypted only, which confirms that encrypting DC alone still leaves some contour information unprotected. The two middle curves are the results involving MV encryption for inter blocks and AC encryption for intra blocks, where the ESS scores are low at the beginning of a GOP and increase substantially toward the end of the GOP. This is because as it approaches the end of a GOP, motion compensation becomes less effective and the compensation residue provides a significant amount of edge information. The information leakage by encrypting DC and MV only has also been seen in Fig. 1. On the other hand, by incorporating the shuffling of macroblock coding units, the resulting ESS measurements are consistently around 0.1 or lower.

The relative overhead of the encrypted videos are listed in Table 2. In general, fast motion videos will have a smaller relative overhead. Since coded unit shuffling does not introduce bitrate overhead, the overhead under encryption settings E4-E6 are identical to that of settings E1-E3, respectively. For a better tradeoff between security and bitrate overhead, we found that setting E5 is suitable for many applications. Reference [9] provides more details.

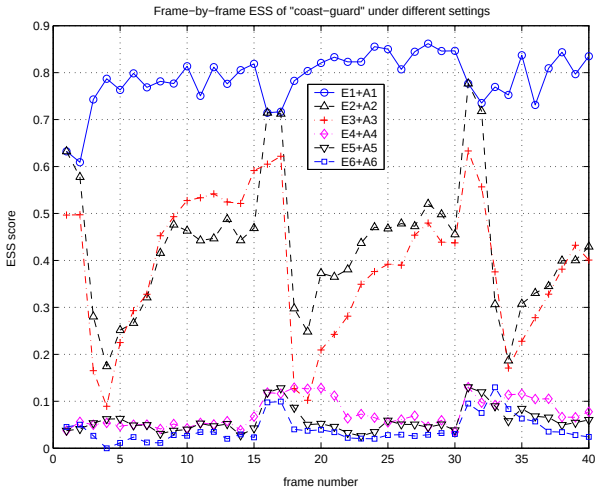

Fig. 2. Frame-by-frame ESS of *coastguard*

4.2. Protecting FGS Enhancement Layer Video

We use 10 frames from the *Foreman* to demonstrate the protection of the enhancement layer while preserving the FGS characteristics. The proposed intra bitplane shuffling is applied within each 8x8 block. We also encrypt the sign bit of each coefficient using a stream cipher. Two encryption settings are used: (a) to shuffle the 1st FGS bitplane, and (b) to shuffle the first two bitplanes. To focus on the protection of the enhancement data, the encrypted FGS bitplanes are combined with a cleartext base layer video to show the visual effects of encryption.

Fig. 3. Encryption results for *Foreman* FGS video. Top row, left to right: base layer plus 1 and 2 unencrypted FGS bitplanes; Bottom row, left to right: encryption settings (a) and (b).

Fig. 3 shows the unencrypted and the encrypted versions of the *Foreman* FGS video, and Table 3 lists the corresponding average PSNR, LSS and ESS. From these results we can see that, without encryption, the ESS, LSS, and PSNR increase with the addition of more bit-planes. With encryption, the edge and luminance similarity remains imperfect. This can be explained by viewing the encrypted FGS bitplanes as random noise added to the base-layer video. Since the ESS score is designed to be resilient to noise, the added noise does not affect the ESS score substantially. However, the LSS score in Table 3 caputures the luminance degradation under encryption settings (a) and (b), as can be seen in Fig. 3. Overall, the results indicate that the video quality after encryption is almost the same as that of the base-layer video and much lower than the cleartext base-plus-enhancement video. Thus the premium quality version of the content can be encrypted in a FGS compatible way and discretionarily protected.

Table 3. Intra Bitplane Shuffling

	Base	+1BP	+2BP	(a)	(b)
PSNR	28.8	29.0	33.4	28.59	27.39
ESS	0.85	0.85	0.92	0.85	0.85
LSS	0.28	0.38	0.79	0.28	0.28

5. CONCLUSIONS

In this paper, we have addressed the importance and feasibility of incorporating signal processing into multimedia encryption. Regarding the security metrics, we pointed out the need of quantifying the security against approximation attacks that are unique to multimedia, and proposed a set of multimedia-specific security metrics to complement those for generic data. Using video as an example, we presented a systematic study on how to integrate different atomic operations together to build a video encryption system. Our experiment shows that by strategically integrating selective value encryption, intra-bitplane shuffling, and spatial permutation, the resulting scheme can achieve a good tradeoff among security, bitrate overhead, and compatibility to signal processing.

6. REFERENCES

[1] L. Qiao and K. Nahrstedt: "Comparison of MPEG Encryption Algorithms", *Inter. Journal on Computers & Graphics*, Permagon Publisher, vol. 22, no. 3, 1998.

[2] Y. Wang, S. Wenger, J. Wen, and A. Katasggelos: "Error Resilient Video Coding Techniques", *IEEE Signal Processing Magazine*, vol.14, no.4, pp61-82, July, 2000.

[3] J. Wen, M. Severa, W. Zeng, M.H. Luttrell and W. Jin: "A Format-Compliant Configurable Encryption Framework for Access Control of Video", *IEEE Trans. on CSVT*, vol.12, no.6, pp545-557, June 2002.

[4] T-L. Wu and S.F. Wu: "Selective Encryption and Watermarking of MPEG Video", *Inter. Conf. on Image Science, Systems, and Technology(CISST'97)*, Las Vegas, NV, 1997.

[5] M. Wu and Y. Mao: "Communication-Friendly Encryption of Multimedia", *Proc. of IEEE MMSP'02*, Dec. 2002.

[6] M. Bellare: "Practice-Oriented Provable Security", *Proc. of First Inter. Workshop on Information Security(ISW'97)*.

[7] Z. Wang, L. Lu and A.C. Bovik: "Video Quality Assessment Based on Structural Distortion Measurement", *Signal Processing: Image Communications*, vol.19, no.1, Jan., 2004.

[8] A. Pal, K. Shanmugasundaram and N. Memon : "Automated reassembly of fragmented images", *Proc. of Inter. Conference on Multimedia and Expo*, Baltimore, Jul. 2003.

[9] M. Wu and Y. Mao: "A Joint Signal Processing and Cryptographic Approach to Multimedia Encryption", submitted to *IEEE Trans. on Signal Processing*, Nov. 2003.