

Fast Compression Domain Video Encryption Scheme for H.264/AVC Streaming

Niansheng Mao, Li Zhuo, Jing Zhang, Xiaoguang Li

Signal and Information Processing Laboratory, Beijing University of Technology, Beijing, China

mns150@sohu.com, zhuoli@bjut.edu.cn, zhj@bjut.edu.cn, lxx@bjut.edu.cn

Abstract—This paper presents a novel method for protecting H.264/AVC streaming. The issues on the compression domain video encryption have been pointed out and fully addressed. To optimize the tradeoff between security level and computational complexity, only the most significant bits for reconstructed video quality in H.264/AVC streaming are extracted and encrypted, including the codewords of intra-prediction mode, the low frequency DCT coefficient of intra-frames and motion vector difference (MVD). To keep the encrypted streaming format-compliance to the H.264/AVC standard decoder, the codewords of intra-prediction mode are encrypted with IPME, the sign bit of the codewords of DCT coefficient are encrypted and the info_suffix of the codewords of MVD are encrypted. Experimental results show that the proposed scheme exhibits significant computational efficiency and reliable security, can resist not only perceptual attacks but also brute-force attacks. Thus, the proposed scheme will be well suited for real-time and resource-limited systems.

Index Terms—Video encryption, H.264/AVC streaming, compressed domain.

I. INTRODUCTION

With rapid development of multimedia communication systems and networks, security and privacy are becoming more and more demanding for images and videos. Such sensitive data needs to be protected before transmission or storage. One approach for protecting multimedia content is to use authentication control mechanisms. However, it cannot sufficiently ensure the physical security for multimedia content. Another straightforward approach is to encrypt the entire compressed bitstream with conventional cryptographic algorithms such as DES [1] or AES [2]. Unfortunately, multimedia content is often of huge volume and requires real-time operations. Meanwhile, in the case of the wireless mobile terminals, limited processing power, memory and bandwidth always fail to meet the encryption processing overhead. Thus, fast video encryption schemes need to be carefully designed.

In real-world applications, a video encryption scheme should take various requirements into account, such as security, computational efficiency, compression efficiency, format-compliance and so forth. Different video applications require variable levels of security. For example, for Video on Demand (VoD) or pay-TV, low security is often required,

and even nonpaying users are allowed to access low quality versions to promote them to buy high quality versions, whereas for military secrets or financial information, strict security is demanded to completely prevent the unauthorized access. The computational efficiency means that the encryption or decryption process can not cause too much time delay, in order to meet the requirements of real-time applications. Video compression is employed to reduce the storage space and save bandwidth, so that the encryption process should have a least impact on the compression efficiency. The format-compliance, also known as syntax-compliance [3], means that the encryption scheme should do not change the syntax structure of the compressed bitstream, thus ensures features like cutting, copying, adding or removing, and ability of the encrypted bitstream still can be decoded by a standard decoder.

In the past decades, a number of video encryption algorithms have been proposed. According to their association with video compression, these algorithms can be classified into joint compression and encryption algorithms and compressed domain encryption algorithms [4]. For the former one, the encryption algorithms are embedded in a certain step of the video compression process. For example, some algorithms permute or scramble the DCT coefficients after the Discrete Cosine Transform (DCT) [5]-[7], some algorithms encrypt the signs of DCT coefficients or motion vector difference (MVD) after quantization [8], [9], and some algorithms selectively encrypt intra-prediction modes, DCT coefficients, and MVDs during the entropy coding [10]-[12]. These algorithms maintain the format-compliance after encryption, but all of them require a modified structure of the standard video codec, leaving all existing codec unusable. In addition, the compression efficiency is usually degraded by permuting or scrambling the DCT coefficients during the coding process.

Differently, for the compressed domain encryption, the compression and encryption process are carried out separately. In [13], the odd indexed bytes in video streaming are firstly encrypted with a conventional cryptographic algorithm, and used as keys to XOR with the even indexed bytes. In [14], the streaming according to their importance for decoding are divided into five types, and the first three

are encrypted whereas others remain. Both of the above algorithms can ensure enough security, but are all of low computational efficiency and loss the format-compliance. In [3], the codewords of DCT coefficients and MVD in compressed bitstream are shuffled. In [15], the codewords of intra-prediction mode are encrypted. Both of them demonstrate high computational efficiency and maintain the format-compliance, but are of low security. As can be seen, the various existing compressed domain encryption schemes cannot optimize the tradeoff between security level and computational complexity, and are often difficult to maintain the format-compliance. Therefore, the compressed domain video encryption algorithms need to be further studied.

In this paper, we propose a fast compressed domain video encryption scheme for H.264/AVC streaming. The proposed scheme directly extracts the most significant bits for video reconstruction in H.264/AVC streaming, concatenates them in an appropriate way to form a sub-bitstream, encrypts the sub-bitstream with a conventional cryptographic algorithm such as AES, and then puts the encrypted bits back into their original positions.

The rest of this paper is organized as follows. Section II analyzes the H.264 bitstream syntax structure. Details of the proposed video encryption scheme are described in Section III. The performance of the proposed scheme is discussed in Section IV. Section V presents some conclusions.

II. H.264 BITSTREAM SYNTAX STRUCTURE

H.264/AVC is the state-of-the-art video coding standard of ITU-T and ISO/IEC [16]. In general, the basic unit of the H.264 bitstream is VLC or FLC codewords, which are formed by a number of bits and represent different information types. These codewords play different roles in the decoding process. For example, the codewords of the header include synchronization information, and the codewords of MVD contain video motion information and so forth.

To achieve higher compression efficiency, the H.264 bitstream is organized with a hierarchical structure, as shown in Fig. 1. The H.264 bitstream can be divided into a series of Network Abstraction Layer (NAL) units. Each NAL unit contains NAL header information and a Raw Byte Sequence Payload (RBSP), which can be Sequence Parameter Set (SPS), Picture Parameter Set (PPS) or a coded slice. Among them, SPS contains SPS_id, profile and level, the number of reference frames, image width and height, and so on. PPS contains PPS_id, SPS_id, entropy coding mode, reference frame index, the initial quantization parameter (QP), and so on. The SPS and PPS do not correspond to a particular sequence or image, in other words, an SPS can be used for multiple sequences and a PPS can also be used for multiple images. The coded slice consists of slice header (including slice type, PPS_id and QP offset) and a number of macroblock (MB) data.

For video encryption, the SPS, PPS and slice header only

provide nominal security, as these coding parameters do not contain too much information and usually have a fixed format. The MB data can be classed into intra-macroblock data (MB_type, intra-prediction mode, coded block pattern (CBP) and DCT coefficients) and inter-macroblock data (MB_type, inter-prediction mode, CBP, MVD and DCT coefficients). Among them, the DCT coefficients contain video texture information, the MVD contain video dynamic information, and the intra-prediction mode indicates the predicted direction. All these information are the most important for video reconstruction. Therefore, in order to obtain high security, the codewords of intra-prediction mode, DCT coefficient, and MVD in H.264 bitstream should be encrypted.

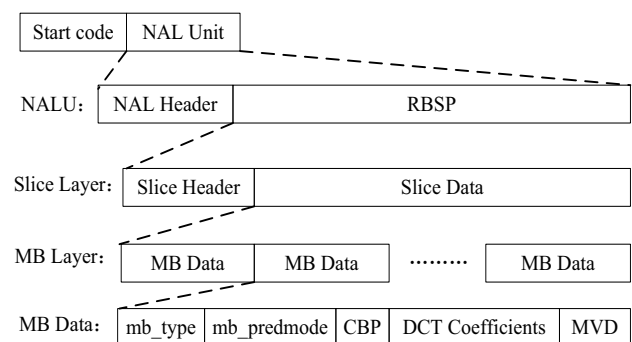


Fig. 1. H.264 bitstream hierarchical structure.

III. THE PROPOSED ENCRYPTION SCHEME

Based on the above analysis, a fast H.264 compressed bitstream encryption scheme is proposed in this section. For intra-prediction mode, the codewords are encrypted with IPME algorithm [15], for the low frequency DCT coefficients of intra-frame, the sign bit of the codewords are encrypted, and for motion vector difference, the info_suffix of the codewords are encrypted. The proposed scheme is shown in Fig. 2.

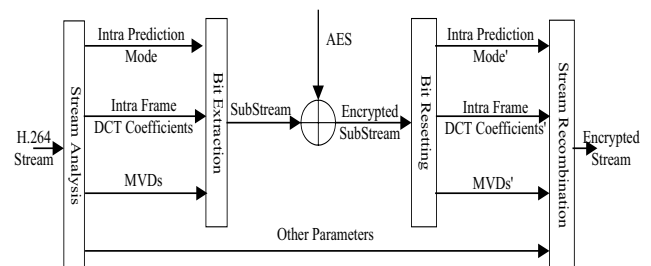


Fig. 2. Diagram of the proposed encryption scheme.

A. Intra-Prediction Mode Encryption

In H.264 bitstream, the intra4×4 prediction mode is denoted by two syntax elements, prev_intra4×4_pred_mode with 1bit and rem_intra4×4_pred_mode with 3 bits. If the prev_intra4×4_pred_mode is set to '1', the current block uses the most_probable_mode, which is the minimum of the

prediction modes of its two neighboring upper and left blocks, and the `rem_intra4×4_pred_mode` is not adopted. Otherwise, the `prev_intra4×4_pred_mode` is set to '0', and the prediction mode of the current block is presented by the `rem_intra4×4_pred_mode`.

In IPME algorithm [15], only the codewords of the `rem_intra4×4_pred_mode` in H.264 bitstream are encrypted. The algorithm is simple and computationally efficient, but lacks of security due to the limited encryption space. In the improved algorithm [17], in order to obtain higher security, when the `prev_intra4×4_pred_mode` is set to '1', the encryption operation is to reset it to '0' and inserts 3 bits chaotic sequence as the `rem_intra4×4_pred_mode`, and all the `intra4×4` prediction modes are encrypted by chaotic pseudo random sequence. However, the improved algorithm significantly increases the computational complexity, and bears a large amount of processing overhead and memory requirements. Thus, the codewords of `intra4×4` prediction modes are encrypted with IPME in the proposed scheme.

Differently, the `intra16×16` prediction modes are jointly encoded with the luma and chroma CBP using the unsigned Exp-Golomb entropy coding. The CBP indicates which blocks within a macroblock contain DCT coefficients, so that its values should not be changed during encryption, otherwise the encrypted bitstream will loss the format-compliance. Thus, the proposed scheme does not encrypt the `intra16×16` codewords.

B. DCT Coefficients Encryption

The DCT coefficients in H.264 baseline profile are encoded with the context-based adaptive variable length coding (CAVLC). The encoding process can be described as follows [16]:

- Encoding the number of coefficients and trailing ones (`coeff_token`),
- Encoding the sign of each trailing ones,
- Encoding the levels of the remaining non-zero coefficients,
- Encoding the total number of zeros before the last coefficient, and
- Encoding each run of zeros.

Since it is context adaptive, the choice of VLC table to encode each level adapts to the magnitude of each successive coded level. Therefore, in order to maintain the format-compliance, the context adaptive property should not be destroyed. In other words, the magnitude of coded levels should not be changed. Thus, only the sign bits of coded level in H.264 bitstream can be encrypted. For optimize the tradeoff between security and computational complexity, only the sign bit of the codewords of the low frequency DCT coefficients of intra-frame are encrypted in the proposed scheme.

C. MVD Encryption

In H.264, each MVD is independently coded by the

signed Exp-Golomb entropy coding. It means that the MVD codewords in H.264 bitstream are mutually independent. Furthermore, each MVD codeword in H.264 bitstream is constructed as $[M \text{ Zeros}][1][\text{INFO}]$, where INFO is a M-bit suffix information called `info_suffix`. Here, the MVD level is $X = 2^M + \text{INFO} - 1$ and the last one bit of the `info_suffix` is the MVD sign. Therefore, the entire `info_suffix` of MVD codewords are extracted and encrypted in the proposed scheme.

IV. PERFORMANCE ANALYSIS

In our experiments, a variety of standard video sequences such as "Akiyo", "Foreman", "Mobile", "News", "Football", "Mother & daughter", "Tempete" and "Silent" are applied to demonstrate the performance of the proposed encryption scheme. Each video sequence with CIF or QCIF format consist of 150 frames. These videos are all encoded by JM86 with a frame rate 15Hz, and the intra-frame period is set as 15. The proposed scheme's performance, such as security, computational complexity and memory requirement, is analyzed as follow.

A. Security

1) Perceptual security

If the videos encrypted with an encryption algorithm are too chaotic to be understood, then the encryption algorithm is regarded as of high perceptual security [11]. Fig. 3 shows several encrypted frames of "Football", "Foreman" and "Mobile" videos using the proposed encryption scheme. As can be seen, by encrypting the most critical bits for video reconstruction in H.264 bitstream, the encrypted videos appear unrecognizable. Furthermore, the three encrypted videos quality expressed in PSNR are illustrated in Fig. 4. As can be seen, the PSNR value of the encrypted video is much lower than the corresponding unencrypted video, and the difference is about 25 dB. Thus, the proposed scheme can achieve a high perceptual security.

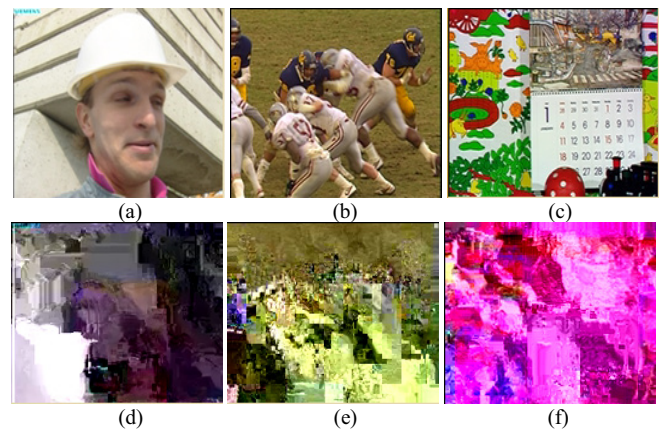


Fig. 3. The encrypted results of the proposed encryption scheme. (a)(b)(c) are the original videos, (d)(e)(f) are the encrypted videos for (a)(b)(c), respectively.

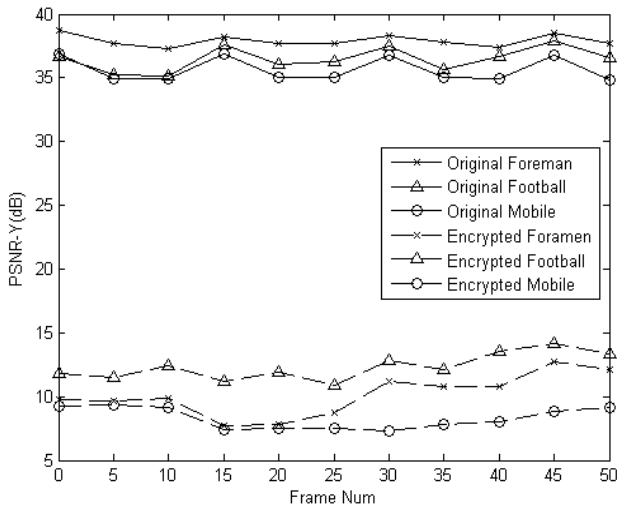


Fig. 4. The video quality in PSNR (dB) versus frame number, for the original videos and the encrypted videos.

2) Perceptual attacks

The video encryption algorithms have to withstand not only classical cryptanalytic attacks such as known-plaintext attacks, chosen-plaintext attacks and ciphertext attacks, but also the perceptual attacks such as error-concealment-based attacks [3] and replacement attacks [10]. The former means that attackers treat the encrypted parts as the loss caused by bit-error or packet-loss and recover them with some error-concealment techniques. The replacement attack is that attackers replace the encrypted parts with arbitrary data and make the encrypted video recognizable. Fig. 5 shows the encrypted videos recovered by error-concealment-based attacks and replacement attacks. As can be seen, all of the recovered videos are still unrecognizable. Thus, the proposed scheme can be against the perceptual attacks.



Fig. 5. The results of perceptual attacks. (a) is the error-concealment-based attack result, (b) is the attack result by replacing the intra-prediction mode, (c) is the attack result by replacing the DCs.

3) Brute-force attacks

Since the AES key lengths support 128, 192 and 256 bits, the minimum brute-force space for each slice is 2^{128} , which is too large for attackers to break. Thus, some attackers try to recover intra-prediction modes, DCT coefficients and MVDs' sign bits with brute-force attacks. In the proposed scheme, the brute-force space of intra prediction mode, the suffix of the codewords of DCs, and the sign bit of ACs and MVDs is 2^3 , 2^R (R is the length of the suffix of codewords of DCs) and 2^2 , respectively. Therefore, for a slice with M

intra-macroblocks and N inter-macroblocks, the brute-force space is $S = [2^3 \cdot 2^R \cdot (2^2)^3]^M \cdot [2^R \cdot 2^2]^N = 2^{(R+9)M} \cdot 2^{(R+2)N}$. By taking QCIF (176×144) video sequences for example, for intra-frame, $M = (176 \times 144) / (16 \times 16) \times 50\% \approx 50$, $N = 0$, the brute-force space is $S \geq 2^{450}$, and for inter-frame, $M = 0$, $N = (176 \times 144) / (16 \times 16) = 99$, the brute-force space is $S \geq 2^{198}$. Thus, the brute-force space is also large enough to withstand attacks.

B. Computational complexity

The computational complexity of the proposed scheme depends on the extraction and encryption of the codewords of intra-prediction modes, DCT coefficients and MVDs. As can be seen from section II, according to the syntax structure, slice header and macroblock header for each slices and each macroblocks in H.264 bitstream can be quickly located. Thus, the extraction process does not take much time. Furthermore, due to AES can be operated as a stream cipher in the OFB mode, the sub-bitstream encryption process also does not cost much time.

The computational efficiency of the proposed scheme (called PVEA) and the SEAH264 algorithm [11] are given in Table I. As can be seen, for the proposed scheme, the time ratio of encryption/compression is no more than 1%, and the time ratio of decryption/decompression is only about 5%. Moreover, the proposed scheme is more time efficiency than the SEAH264 algorithm. Thus, the proposed scheme has significant computational efficiency.

TABLE I
COMPARISON ON COMPUTATIONAL EFFICIENCY

Video	Size	Time ratio			
		Encryption/Compression		Decryption/Decompression	
		SEAH264	PVEA	SEAH264	PVEA
Foreman	QCIF	0.9%	0.5%	5.2%	2.9%
Akiyo	QCIF	1.1%	0.6%	4.9%	4.6%
Mother	QCIF	0.7%	0.3%	5.9%	2.8%
News	QCIF	0.5%	0.2%	4.7%	3.9%
Akiyo	CIF	0.7%	0.4%	6.1%	3.6%
Foreman	CIF	1.0%	0.3%	6.2%	4.3%
Football	CIF	1.3%	0.6%	4.8%	5.2%
Mobile	CIF	0.9%	0.8%	6.2%	5.8%

C. Memory requirement

Generally, the larger amount of data is processed in encryption, the more memory is required. Therefore, the memory requirement of the proposed scheme can be measured by the amount of the encrypted data. Fig. 6 shows the ratio between the encrypted data and the corresponding slice data. As can be seen, for intra-frames, the encrypted data is about 15% of the corresponding slice data, and for inter-frames, the ratio is only about 5%. That is because, considering intra-frames are more important than inter-frames, the proposed scheme provides enhanced encryption

to this kind of information. All in all, compared with the existing encryption schemes, the proposed scheme adds less overhead of memory requirement.

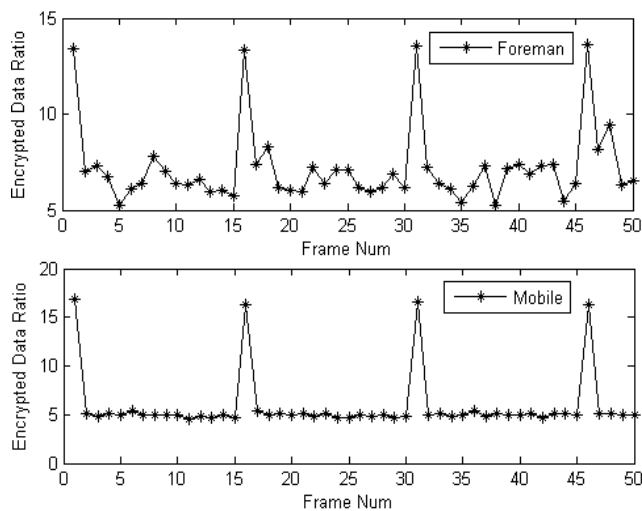


Fig. 6. The ratio between the encrypted data and the slice data versus frame number.

V. CONCLUSION

In this paper, a novel video encryption scheme in H.264 compressed domain is proposed, which is independent of the compression process, thus does not need to modify the structure of a standard video codec. In order to reduce the processing overhead and memory requirement, the proposed scheme only encrypts the `rem_intra4x4_pred_mode` of the codewords of intra-prediction modes, the sign bit of the codewords of the low frequency DCT coefficients of intra-frame, and the `info_suffix` of the codewords of MVD in H.264 bitstream. By optimizing the trade-off between security and computational complexity, the proposed scheme exhibits reliable security, can resist not only perceptual attacks but also brute-force attacks, and obtains significant computational efficiency. Furthermore, it maintains the format-compliance to the H.264 standard decoder and has no escalation in bitrate. Thus, the proposed scheme is better suited for real-time and resource-limited systems.

ACKNOWLEDGMENT

This research is supported by the Program for New Century Excellent Talents in University, the National Natural Science Foundation of China (No.61003289, No.61100212), the Natural Science Foundation of Beijing (No. 4102008), the Excellent Science Program for the Returned Overseas Chinese Scholars of Ministry of Human Resources and Social Security of China, Scientific Research Foundation for the Returned Overseas Chinese Scholars of MOE and the Scientific Research Foundation of Beijing University of Technology for Doctor/Master (No. ykj-2011-5474).

REFERENCES

- [1] Data Encryption Standard (DES), FIPS PUB 46, Jan. 1977.
- [2] Advanced Encryption Standard (AES), FIPS-PUB 197, Nov. 2001.
- [3] J. Wen, M. Severa, W. Zeng, *et al*, "A format-compliant configurable encryption framework for access control of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 545-557, Jun. 2002.
- [4] F. Liu and H. Koenig, "A survey of video encryption algorithms," *Computers and Security*, vol. 29, pp. 3-15, Feb. 2010.
- [5] L. Tang, "Methods for encrypting and decrypting MPEG video data efficiently," *Proceedings of the 4th Multimedia Conference (ACM Multimedia 96)*, pp. 219-229, Boston, MA, USA, Nov. 1996.
- [6] A. S. Tang and W. C. Feng, "Efficient multi-layer coding and encryption of MPEG video streams," *IEEE International Conference on Multimedia and Expo.*, vol. 1, pp. 119-122, Aug. 2000.
- [7] W. Zeng and S. Lei, "Efficient frequency domain selective scrambling of digital video," *IEEE Transactions on Multimedia*, vol. 5, no. 1, pp. 118-129, Mar. 2003.
- [8] C. Shi and B. Bhargava, "An Efficient MPEG video encryption algorithm," *Proceedings of the 17th IEEE Symposium on Reliable Distributed Systems*, pp. 381-386, Oct. 1998.
- [9] C. Shi, S. Wang, and B. Bhargava, "MPEG video encryption in real-time using secret key cryptography," *In Proc. of PDPTA'99*, Las Vegas, Nevada, pp.2822-2828, 1999.
- [10] C. P. Wu and C. J. Kuo, "Design of integrated multimedia compression and encryption systems," *IEEE Transactions on Multimedia*, vol. 7, no. 5, pp. 828-839, Oct. 2005.
- [11] S. Lian, Z. Liu, Z. Ren and H. Wang, "Secure advanced video coding based on selective encryption algorithms," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 2, pp. 621-629, May 2006.
- [12] J. Zhou, Z. Liang, Y. Chen and O. C. Au, "Security analysis of multimedia encryption schemes based on multiple Huffman table," *IEEE Signal Processing Letters*, vol. 14, no. 3, pp. 201-204, Mar. 2007.
- [13] L. Qao and K. Nahrstedt, "A new algorithm for MPEG video encryption," *Proceedings of the First International Conference on Imaging Science, Systems and Technology (CISST'97)*, pp. 21-29, Las Vegas, Nevada, July 1997.
- [14] T. Shi, B. King and P. Salama, "Selective encryption for H.264/AVC video coding," *In SPIE International Society for Optical Engineering (San Jose, CA, USA)*, vol. 6072, pp. 171-179, Feb. 2006.
- [15] J. Ahn, H. Shim, B. Jeon and I. Choi, "Digital Video Scrambling Method Using Intra Prediction Mode," *Advanced in Multimedia Information Processing PCM2004*, vol. 3333, pp. 386-393, Dec. 2004.
- [16] T. Wiegand, G. J. Sullivan, G. Bjntegaard, *et al*, "Overview of the H.264/AVC video coding standard," *IEEE Tran. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, July 2003.
- [17] J. Jiang, S. Xing and M. Qi, "An intra prediction mode-based video encryption algorithm in H.264," *International Conference on Multimedia Information Networking and Security (MINES 2009)*, pp. 478-482, Hubei, China, Nov. 2009.