

# A Scalable Security Service Model for Reliable End-to-End Media Distribution and Reuse

Yong-Hyuk Moon<sup>1</sup>, Hyeokchan Kwon<sup>1</sup>, Jae-Hoon Nah<sup>1</sup>, and Chan-Hyun Youn<sup>2</sup>

<sup>1</sup> Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea

<sup>2</sup> Korean Institute of Science and Technology (KAIST), Daejeon, South Korea

{yhmoon, hckwon, jhnah}@etri.re.kr; chyoun@kaist.ac.kr

**Abstract.** Due to lack of considerations on media scalability and reuse, most selective encryption schemes previously proposed might be hard to meet challenging requirements with a certain level of security which are induced from the multi-layer based video technique, for example scalable video coding. As a reason of that, first we discuss security concerns on this issue and then analyze the existing selective encryption algorithms in a media scalability point of view. Secondly, we propose the dynamic load balancing strategy for re-adjusting protection strength of reconstructed media and suggest operational mode which can be more suitable for media reuse. Finally, we demonstrate scalable security mechanism's decryption overhead compared with decoding burden in time and effect of visual distortion through our designed simulation.

**Keywords:** Video Streaming Security, Scalable Security, Media Scalability, Scalable Video Coding

## 1 Introduction

Recently, a video streaming technology [1, 2, 3] which has huge popularity in IPTV, Internet TV, and web streaming with users' and content providers' generated contents using Internet-based CDN or P2P-style networking techniques becomes a major industry; indeed, the video is a king of the multimedia.

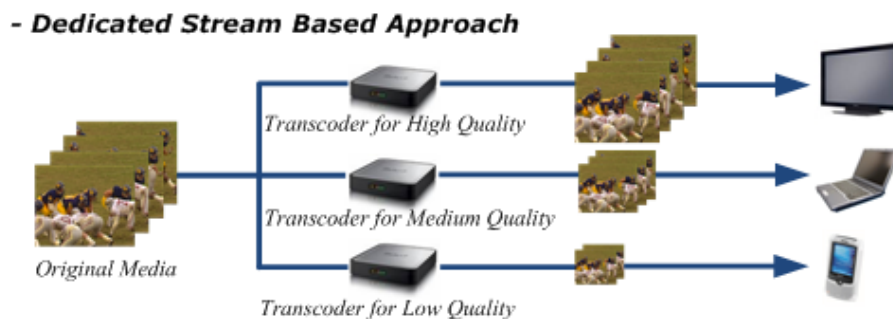
Infrastructure related researches focusing on a networking structure, scheduling mechanism, system reliability and capacity analysis, and deployment issues, have been actively studied until now. However, vulnerabilities and security holes in streaming media are not considerably treated as main problems in a research area of video distribution. Environments of media distribution are changing to more interactive, multi-directional, and reusable in usage patterns. Thereby, media can be freely delivered to anyone (even unauthorized) at anytime, to any places, and with any quality of video. In other words, media scalability should be considered with distribution and reuse among many end points. Besides, media technique (frequently referred to as video compression or codec) has been developed for providing more flexible scalability with high coding efficiency (scalable media); thus, a new security mechanism for scalable media (e.g., scalable video coding [4, 5]) is truly required.

However, most proposed algorithms tended to mainly deal with a single layer video coding such as MPEG4 [1], H.264/AVC [5, 6] as a media format. In fact, multi-layer encoded video files more certainly might be exposed to unexpected harms and outflow of original data than strict-format media; since, scalable media could be transformed from one source to multiple media streams by reconstruction process according to required use cases.

In general, media diversity regarding scalability comes with more complicated concerns on security. In this paper, therefore, we discuss security considerations to redesign existing protection schemes for supporting reliable end-to-end scalable media distribution and reuse with respect to protection strength and burden of cryptographic operation including encryption and decryption.

## 2 Security Concerns in Media Scalability

For supporting media scalability, a transcoding (downscaling) technique was conventionally used. In this approach, dedicated stream is required for each service requirement (e.g., a device's capacity and a demanded service level) as shown in Fig. 1. Additionally, this method assumes that a transcoder must know all information which is related to encoding and decoding; because, in order to manipulate high quality of video stream to low one, the transcoder should decode and then encode original data again for each scenario [7]. As a reason of this procedure, it is intuitively expected that the transcoder is venerable to a single point of failure for content extraction due to an illegal trial of forge, DoS attack, and unauthorized access.



**Fig. 1.** Transcoding Service Model

For secure transmission, in the early stage of applying streaming security, encrypting the entire bit-streams was simply used; however, a fact that decrypting, decoding, and re-encoding processes are necessary at the transcoder was tuned to be a critical drawback. Accordingly, it was expected that original video data is inevitably exposed in the procedure of transcoding video streams such as MPEG and H.264. Therefore, without supporting new video compression standard, a secure downscaling technique, which does not need the decryption, might not possibly be archived.

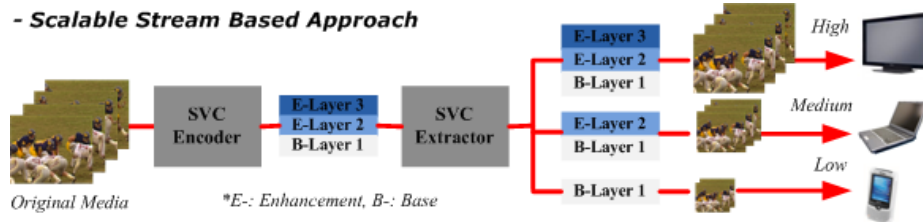
Currently, the SVC (Scalable Video Coding) standard [4, 5] based on the H.264/MPEG-4 AVC video compression technique becomes obtaining more

popularity in video streaming related research areas and industries due to its scalable features with high coding efficiency. In particular, the SVC constructs scalable bit-streams consisting of a base layer and enhancement layer and this design can hide the detailed structure of video by using NAL (Network Abstract Layer) which allows for high level readability of stream packets. Through this structural characteristic, it can offer three kinds of scalabilities in resolution, frame rate, and bit rate for heterogeneous devices by extracting original bit-streams to required layers in order to reconstruct appropriate bit-streams; thus, it is very suitable for different use cases as depicted in Fig. 2.

Basically, this unique structure of SVC can be formulated as Eq. (1).

$$V = \sum_{i=1}^m c_i = \sum_{i=1}^m (H_i + B_i + \sum_j^l E_j) \quad (1)$$

Where, SVC-encoded video chunks ( $c$ ) usually consist of headers ( $H$ ), including signaling information; base layers ( $B$ ), exactly same as the H.264 single layer; and enhancement layers ( $E$ ), with high quality of bit-streams.



**Fig. 2.** SVC Service Model

In this paper, therefore, the media scalability can be defined as any layer-encoded media using the SVC standard; this media is variable and flexible in terms of temporal rate, spatial size, and video quality, under certain conditions which are time limitation (e.g., deadline), network bandwidth, packet loss rate, and so forth.

Next, we will discuss the scalable security for media (re-)distribution in section 3.

### 3 Scalable Security for Media Distribution and Reuse

The SVC media can be distributed to receivers (user's playing devices) in the form of NAL packets with different scalability according to service requirements and device capacities. The distribution path of SVC media is not unidirectional and can be formed as the full-mesh sharing; thus, any unauthorized access to path during media redistribution or reuse among different devices (users) possibly induces security problems.

Hence, in this section we evolved new security mechanism for the purpose of re-design with the concepts of existing selective encryption schemes for MPEG, H.264, and SVC. We then expected that through the evolved considerations on this issue [8]

the scalable security service model can avoid the derivative vulnerabilities occurred in media distribution and reuse among users.

### 3.1 Security Mechanism for Scalable Media

Some researchers did not tend to concern about format-compliance [9, 10] when they encrypt frames such as I (Intra)-frame, P (Predictive)-frame, and B (Bi-directional Predictive)-frame. This trial of manipulating bit-streams is intuitively to induce unpredictable errors when decoding; because, arbitrarily modified bit-streams might not be compliant to the video compression standard including very complex rules and algorithms such as discrete cosine transform (DCT), variable length coding (e.g., CAVLC, CABAC), intra frame prediction, and motion compensation/estimation. In other words, prototype (standard type) or experimental purpose decoder might not provide the wide compatibility as one of commercial version which is re-designed to a particular application.

In opposite to these contributions above, others tried to encrypt signaling information located in a field of header to avoid format compliance issue due to a strict decoder as well as to reduce the encryption overhead considered as one performance index. Unfortunately, cryptographic information in the form of additional application-level headers was very necessary since a decrypting module at a receiver needs to comprehend which parts are encrypted, what encryption algorithm is used, what blank data is added as a pending segment, and so on for correct interpretation before performing the decryption.

These selective encryption algorithms were proposed in order to reduce encryption load and eliminate unnecessary or additional management information. However, it is not directly appropriate to the scalable media supported by SVC design; since most schemes were devised for the single layer media encoded by MPEG and H.264 only as a target media format.

**Classification and Comparison of Scalable Media Security.** Selective encryption mechanisms can be widely classified to three classes; it is dependent on a unit of encryption object such as 1) layers, 2) frames, and 3) signaling information. However, it can cover most proposed schemes. Moreover, each unit can be encrypted in the combinational form depending on service scenarios.

*1) Layer Level Encryption.* The most straight-forward method to obtain secure transmission of scalable media could be to encrypt whole part of SVC bit-streams, for example entire parts of base layer. The base layer contains the lowest level of quality information of video which can be independently decodable, while enhancement layers cannot be decoded without the base layer. Therefore, a simple conditional access can be achieved using this trait of encrypting the base layer. This seems to be the most secure approach (referred to as wholesale encryption [11]). However, we believed that more intelligent approach will reduce the computational overhead without compromising the required strength of security during transmission.

2) *Frame Level Encryption*. The reference frames are very significant since B frames (as non-reference frame) must see I or P frame within GOP (Group of Picture) in order to reconstruct correct frames. In other words, the decoder basically conforms to the integrity of the reference frames to decode the P or B frames due to the inter-frame dependency of SVC. Hence, the encryption of reference frames offers a certain level of security. Moreover, encrypting I frames and (all or partial) I blocks in the P and B frames increases the amount of invisible information in the media. In SVC, this scheme can be extended to a way that encrypting whole I frames (and/or I blocks) of base layer and enhancement layers. More detailed effect of visual distortion can be found in [11].

3) *Signaling Information Level Encryption*. As the lightest way of encrypting SVC bit-streams with supporting a certain security level, encryption of texture (residual data), motion vector difference value, FGS, and intra prediction modes were proposed in [12, 13, 14], respectively. Each of them can be used in the individual or combinational form. Accordingly, if more values are encrypted, then distortion level of scenes greatly increases. The trade-off relation between encryption performance; regarding time and security level; in terms of distortion rate of video, should be carefully investigated before taking one option of given cryptographically selective algorithms on scalable media.

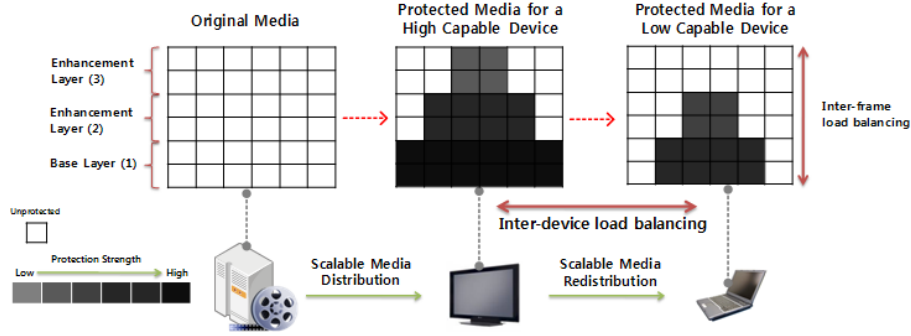
In short, however, these schemes are not perfectly free from such issues including the format-compliance, unpredictable fluctuation of video compression rate, encryption time overhead, decoding performance for real-time streaming service.

On the other hand, one more issue to consider is about protection strength; in fact, media sharing technique might not be actively supported under MPEG-oriented video streaming environment due to lack of considerations on media scalabilities. So, an encryption scheme providing the same protection strength was simply applied to an individual media, regardless of receivers' level of capacity and demand. Namely, load balancing of encryption during media distribution should be significantly reflected. For this, in perspective of media scalability, we re-designed the existing selective encryption algorithms with dynamic protection strength considering load balancing strategy as follows.

**Dynamic Load Balancing Strategy of Scalable Security Service Model.** As shown in Fig. 3, the first type is the inter-layer load balancing; encryption parts are distributed between base layer and enhancement layers for supporting cryptographic fairness (invisibility) and reducing performance overhead. Secondly, when the encrypted scalable media is transmitted to a heterogeneous-capable device, inter-device load balancing can be used for adjusting a level of protection strength.

Previous studies tended to propose that encrypting a base layer is enough to satisfy contents providers' or owners' security requirements. These research results do not always make this assertion widely acceptable and useful, even though this scheme could archive some stable level of security; since it is obviously expected that differential service levels are required by users and heterogeneous-capable devices

participates in media distribution networks. Particularly, in multi-layers encoding rule, parts of encryption could be a base layer as well as enhancement layers; thus, this means that we can flexibly select which parts are effective to encryption load balancing strategy and also guarantee strong protection strength for satisfying user's required service level.



**Fig. 3.** Dynamic load balancing mechanism for reducing encryption overhead

In Fig. 3, the original media from a server to the first consumer (high-capable device) was probably encrypted at strong level of protection. Then this media could be re-encrypted without decrypting, decoding, and re-encoding processes for redistribution to the second device (a wired laptop) with lowering level of security as well as with considering the laptop's capacity. This process will be securely repeated to the next devices. Hence, we can achieve secure media transmission even under the end-to-end redistribution (or reuse) usage cases. However, a decision function should be provided to make a good policy for the redistribution of scalable media; this problem could be abstractly formulated in the form of objective function as follows:

Given algorithm set:  $A_x = \{x_1, x_2, x_3, \dots, x_N\}$

Given encoded scalable media:  $M_s$

Max  $D_e = \sum_{i=1}^s D(s_i, E_{x_j}(s_i))$  for  $\forall j = 1, 2, 3, \dots, N$

s.t.  $Min \frac{T_e(x_j)}{S}$ , (2)

$\frac{R(E_{x_j}(M_s)) - R(M_s)}{R(M_s)} \leq \alpha$ ,

satisfying format-compliance

Where, we supposed that format-compliance was already satisfied.  $D()$  estimates the distortion rate of encrypted media,  $s_i$  is a segment (or unit) of full media, and  $E_{x_j}()$  denotes encryption function using algorithm  $x_j$ . Moreover, we defined  $T_e(x_j)$  as the

encryption timer function with  $x_j$  and  $R(M_s)$  represents video compression(coding efficiency) rate function. Finally,  $N$ ,  $S$ , and  $\alpha$  are the number of combinational cases of encryption algorithms, the number of segments of media, and threshold, respectively.

In this problem, the objective fundamentally was to acquire the maximized distortion rate of encrypted scalable media ( $Max D_e$ ) under given algorithm set  $A_x$  and scalable media  $M_s$ . The first conditional function means that average time for encryption of scalable media should be minimized and next condition indicates that fluctuation (difference) of video compression ratio between SVC-encoded original media and encrypted SVC media.

### 3.2 Security Mechanism for Reusable Media

The reusable media, especially reconstructed video, should be transformed into the appropriate form in terms of resolution, frame rate, and SNR quality. Thus, security mechanism should concern about media's features on reconstruction (referred to as media reformation). For this, two types of encryption scheme are considered here to analyze their feasibility and efficiency of media reuse (or redistribution with the manipulated version of media) between two end points.

**NAL-level En(de)-encryption Scheme.** After encoding media, the encoder produces the NAL streams as an output file; next, for media encryption, the encryptor must have an additional NAL parser which interprets boundary (offset) and comprehends optional information of each field such as PPS, SPS, and SEI of SVC NAL. In this approach, one obvious advantage is to design and implement the security module independently to an encoding module; thus, in particular, when media is extracted to required bit-streams, it can support media re-encryption without decryption. For this reason, the load balancing of portions which will be encrypted can be achieved for adjusting protection strength in this approach. On the other hand, this approach does not fit in real-time broadcasting service, since this en(de)-encryption process performs in the batch processing manner so that it is more proper to file-level media redistribution (e.g., VoD streaming). Therefore, this method is more competent and agile to support secure end-to-end media reuse. However, algorithms using NAL-level en(de)-encryption [15, 16] were rarely proposed in light of supporting media redistribution or reuse.

**Encoding-level En(de)-encryption Scheme.** In opposite to the first scheme, the encryption simultaneously performs during encoding process [12, 13, 14]. Shortly, the encoding process will call encryption related functions when it is needed; thus, encryption promptly operates as a subset of encoding module. Accordingly, it might be much easier than NAL-level method to satisfy format-compliance in this approach, only if the encrypting module complies with the context-adaptive entropy coding rules; namely, it depends on a point where encryption performs in the whole procedure of video compression. Besides, it alleviates efforts of implementing an

NAL parser. Shortly, the encoding-level encryption scheme can be classified into two types such as 1) before DCT, spatial domain; after DCT; frequency domain. Despite of this benefit, the scalable media must be decrypted at a receiver or mediator, before performing media downscaling transcoding from one device to other. Due to its capability of supporting restricted cryptographic re-adjustment, encoding-level en(de-)cryption scheme might be more suitable to the live streaming service; since, consuming time of decryption process during decoding at receivers is significantly considered as the most important fact. Consequently, it is possibly certain that this method hardly offers secure end-to-end media reuse.

With more specific points of view, a comparison between two schemes was summarized in table 1.

**Table 1.** A comparison between two scalable security mechanisms for media reuse.

	<b>NAL-level</b>	<b>Encoding-level</b>
Modularity	Independent Additional parser required	Integrated (as sub-module) Depending on an algorithm
Encryption Overhead	Depending on an algorithm Batch processing	Line by line operation
Feasibility for Reuse with differential service level	Readjust the protection strength by no decryption process Pre-defined parsing needed	Entire (or partial) decryption required More proper to real-time streaming service
Decryption Overhead	If decryption time is tiny, it is no matter. Hard to meet this requirement	
Format Compliance		Depending on a point (Entropy Coding) of performing the encryption
Coding Efficiency	It might not be changed in compression rate	Slightly increasing if encrypting after DCT

As a reason described above, these concerns should be merged into new scalable security service model for SVC; namely, the scalable security would better support the NAL-level encryption for reusable media; however, for general use cases in streaming service requirements, two schemes should be optionally chosen.

#### **4 An En(De-)cryption Simulation Tool for SVC**

For our experiment, we developed an en(de-)cryption simulation tool (as pilot software) based on JSVM (Joint Scalable Video Model) which is used as an experimental-purpose SVC encoder and decoder compliant to ITU-T standard [5]. This experiment was designed for simply estimating decryption overhead compared with decoding time and for testing a degree of invisibility (distortion rate); however, this measurement did not cover all considerations discussed in section 3.

In this section, we basically assumed that the encoding and encryption are an off-line process, while the decoding and decryption process should be performed in real-time. This fact came from general video streaming model such as real-time broadcasting and video on demand service; so, it can be generally understandable.

For measuring decryption overhead compared with decoding performance, we set a configuration as follows: 1) we utilized CBR (Constant Bit Rate) type media; this format is understandable because we excluded networking variations such as bandwidth, loss rate, and jitter in this paper; in fact, this will be considered in further study. 2) AES 128bit cryptographic algorithm was applied for encryption and decryption. 3) 15 layers were constructed with 640x480 as the maximum resolution (VGA) and 30 fps as the maximum frame rate. 4) Decryption keys for media access were generated for each layer. 5) From 10 times simulation with the same configuration, we calculated average consuming time for decoding and decryption.

A goal of this simulation was to evaluate average consuming time for decryption; thus, it was reasonable that encryption could be executed regardless of operation modes such as the NAL-level encryption and encoding-level encryption. As the first result, table 2 indicated that approximately 3.59 % overhead calculated by  $B/(A+B)$  was averagely expected for even full-VCL(Video Coding Layer) encryption. This numerical value could be used as reference of encryption load; since, most selective encryption algorithms encrypted smaller portion of video than the naïve method did, for example layer-level encryption. Therefore, an adaptive use (as the individual or combinational form) of selective encryption(s) will decrease the expected performance burden due to cryptographic operations.

**Table 2.** A comparison between decoding time and decryption time.

Video Length	Decoding Time (A)	Decryption Time (B)
1 frame	0.0268 sec	0.001 sec
1 sec	0.804 sec	0.03 sec
1 min	48.24 sec	1.8 sec
60 mins	2,894.4 sec	108 sec

Secondly, based on our survey, there might not be a SVC rendering application to play bit-streams. Hence, we used the open source based YUV player in order for investigating a degree of exposures of original movie in terms of visibility. Fig. 4 showed that distortion effect of 3 scenes, which was used as experimental-purpose only, respectively when user requested the full quality of video and has decryption keys for only base layer.



**Fig. 4.** Effect of Visual Distortion

As shown in Fig. 4, each frame was not considerably distorted by visual noises, although we applied the (VCL-level) wholesale encryption to the video. Despite of

strong encryption strategy, if the user with an authorized permission to access the only base layer, it was possible that the user could obtain somewhat clear movie; since, the base layer usually contains most information according to SVC rules. This was one critical drawback of base layer based conditional access strategy which could be expected in real SVC streaming service.

## 5 Conclusion and Further Study

As discussed previously, the conventional selective encryption schemes could be extended for protecting more scalable media; however, there are still many design requirements such as format-compliance, unpredictable fluctuation of video compression rate, encryption time overhead, and decoding performance for real-time streaming service.

With the concepts of existing selective encryption schemes for MPEG, H.264, and SVC, our security mechanism was re-designed for supporting reliable media distribution and reuse among users as well as for evolving candidate strategy for found constraints in this paper. First, we discussed necessities of dynamic inter-layer and inter-device load balancing policies as main part of scalable security service model. Secondly, NAL-level and encoding-level en(de-)ryption schemes were compared in perspective of media reuse in order to provide more flexible option to our scalable security service model for SVC media. We particularly focused on showing the decryption's performance burden with full-layer encryption scheme and describing the conditional access issue through visual distortion of a SVC-encoded sample movie in our experiment.

In further study, more simulation results should be offered and then analyzed with various views. Furthermore, it will be also necessary to discuss how much compression rate will be fluctuated with difference encryption schemes. This analysis (performance index) seems to be very required for proving proposed algorithms' real efficiency and feasibility to deployment; because, any encryption method must not impede the coding efficiency.

**Acknowledgments.** This work was supported by the IT R&D program of MKE/KCC/IITA [2008-S-006-01, Development of Open-IPTV (IPTV2.0) Technologies for Wired and Wireless Networks].

## References

1. Iain E. G. Richardson, "H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia", published by John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, ISBN 0-470-84837-5.
2. Patrick Seeling, Frank H.P. Fitzek, and Martin Reisslein, "Video Traces for Network Performance Evaluation: A Comprehensive Overview and Guide on Video Traces and Their Utilization in Networking Research", published by Springer, P.P. Box 17, 30300 AA Dordrecht, The Netherlands, ISBN-13: 978-1-4020-5565-2 (HB).

3. John G. Apostolopoulos, Wai-tian Tan, and Susie J. Wee, "Video Streaming: Concepts, Algorithms, and Systems", HP Laboratories Palo Alto, HPL-2002-260, Sept. 18, 2002.
4. Heiko Schwarz, Detlev Marpe, and Thomas Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 9, Sept. 2007.
5. Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of moving video, ITU-T Recommendation H.264, Nov. 2007.
6. Thomas Wiegand, Gary J. Sullivan, Gisle Bjøntegaard, and Ajay Luthra, "Overview of the H.264/AVC Video Coding Standard", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, July 2003.
7. Susie J. Wee and John G. Apostolopoulos, "Secure Scalable Streaming Enabling Transcoding without Decryption", in Proc. of IEEE International Conference on Image Processing, Thessaloniki, Greece, Oct. 2001.
8. Xiliang Liu and Ahmet M. Eskicioglu, "Selective Encryption of Multimedia Content in Distribution Networks: Challenges and New Directions", CIIT 2003.
9. Thomas Stutz and Andreas Uhl, "Format-compliant Encryption of H.264/AVC and SVC", in Proc. of tenth IEEE International Symposium on Multimedia (ISM'08), 2008.
10. WANG Li-feng, WANG Wen-dong, MA Jian, XIAO Chen, WANG Kong-qiao, "Perceptual Video Encryption Scheme for Mobile Application Based on H.264", The Journal of China Universities of Posts and Telecommunications, Sept. 2008, pp. 73-78.
11. Iskender Agi and Li Gong, "An Empirical Study of Secure MPEG Video Transmission", in Proc. of SNDSS 1996.
12. Tuo Shi, Brian King, and Paul Salama, "Selective Encryption for H.264/AVC Video Coding", in Proc. Of SPIE-IS&T Electronic Imaging, SPIE Vol. 6072, 607217.
13. Su-Wan Park and Sang Uk Shin, "An Efficient Encryption and Key Management Scheme for Layered Access Control of H.264/Scalable Video Coding", IEICE Trans. Inf. & Syst. Vol.E92-D, No.5 May 2009.
14. Yong Geun Won, Tae Meon Bae, and Yong Man Ro, "Scalable Protection and Access Control in Full Scalable Video Coding", IWDW 2006, LNCS 4283, pp. 407-421, 2006.
15. Chunhua Li, Xinxin Zhou, and Yuzhuo Zhong, "NAL Level Encryption for Scalable Video Coding", PCM 2008, LNCS 5353, pp 496-505, 2008.
16. Hendry, Munchurl Kim, Sangjin Hahm, Keunsik Lee, and Keunsoo Park, "A Layered Protection Scheme for Scalable Video Coding", in Conference Proc. Of The Korean Society of Broadcast Engineers, pp. 307-312, Nov. 2006.