情報保護學會論文誌
第20卷 第6號, 2010. 12

차세대 IPTV 서비스를 위한 보안 프레임워크 설계*

이 승 민,† 나 재 훈, 서 동 일
한국전자통신연구원

Design of Security Framework for Next Generation IPTV Services*

Seungmin Lee†, Jae Hoon Nah, Dong-Il Seo
Electronics and Telecommunications Research Institute

요 약

최근 디지털 컨버전스가 가속화되면서 급부상하고 있는 차세대 IPTV 서비스는 디바이스에 구애받지 않고 자유롭게 콘텐츠의 생성과 소비가 가능하여, 전송환경과 디바이스의 특성에 맞는 실시간 서비스와 콘텐츠의 재사용 서비스를 확장성 있게 제공함을 특징으로 한다. 본 논문에서는 이러한 차세대 IPTV 서비스를 제공함에 있어서 요구되는 보안 요구조건과 이를 해결하기 위한 보안 프레임워크를 제안한다. 제안 방법은 기본적으로 SVC (Scalable Video Coding)를 사용하는 단일 메커니즘으로써, 서비스가 제공되는 모든 구간에 대하여 높은 보안성을 보장하며, 동시에 안전한 미디어 적응변환과 동적인 보안 강도 조절이 가능하다는 장점이 있다. 본 논문에서는 현실적인 서비스 시나리오를 바탕으로 제안 방법의 타당성을 입증하였고, 보안 기술 자체만으로도 새로운 비즈니스 기회를 창출 할 수 있는 가능성을 제시하고 있다는 점에서 의의가 있다.

ABSTRACT

With the emergence of increasingly complex networks and diverse user terminals, demand for the next generation IPTV service is rapidly growing. It enables any content to seamlessly be reused on the diverse terminals as well as be broadcasted in real-time through the complex networks. In this paper, a novel security framework is proposed for the real-time and reusable IPTV services. The proposed framework is advantageous over the conventional content protection techniques in easily producing the scalable content with lightweight, perceptual, transcodable, and adjustable security features. It does not only ensure end-to-end security over the entire service range based on a single security mechanism, but also can control a level of security while dynamically transcoding the original content. This approach basically performs selective encryption during and after the compression using scalable video coding. The suitability of the proposed approach is demonstrated through experiments with a practical service scenario. Therefore, it is expected that security technology alone could practically contribute to creating new business opportunities for IPTV services.

Keywords: Content protection, IPTV security, SVC encryption

I. Introduction

The television entertainment industry is currently experiencing a major transformation as broadband subscribers and improvements in compression techniques for digital video content continue to grow. This growth is accelerating the demand for a new generation of technology that allows any content to be watched on any device, anytime and

접수일(2010년 4월 29일), 게재확정일(2010년 9월 30일)
* 본 연구는 지식경제부/방송통신위원회 및 정보통신연구진흥원의 IT 핵심기술개발사업의 일환으로 수행하였음. [2008-S-006-01, "유무선 환경의 개방형 IPTV (IPTV2.0) 기술개발"].
† 주저자 및 교신저자, todtom@etri.re.kr

anywhere. That is, the content is not only broadcasted in real-time through complex networks, but also is seamlessly reused on diverse devices.

This next generation IPTV service explores key challenges associated with successfully managing the technical operation of increasingly complex networks and diverse user terminals. The complex networks have led to growing interest in the development of a video codec that can dynamically adapt to the network architecture and temporal variations in network conditions such as bandwidth and error probability. Furthermore, the diverse devices such as smartphones, handheld personal digital assistants and desktop workstations, each of which has different display resolutions and processing capabilities, may all have access to the same digital media content.

In this paper, a novel security framework is proposed that allows any content to seamlessly be reused as well as be broadcasted in real-time on diverse terminals through complex networks. Notably, the proposed framework is advantageous over the conventional content protection techniques in that it can easily produce the scalable content with lightweight, perceptual, transcodable, and adjustable security features. Moreover, it ensures end-to-end security over the entire service range based on a single security mechanism. The suitability of the framework is demonstrated through experiments with a practical service scenario. Therefore, it is expected that security itself could practically contribute to creating new business models for IPTV services.

This paper is organized as follows. Related work is discussed in Section II. The scalable IPTV service is defined and its security requirements are derived in Section III. Then, the security framework for satisfying the requirements is presented in Section IV. In Section V, a service model is assumed and then experiments are performed under the model. The study concludes in Section VI with a summary and plans for future research.

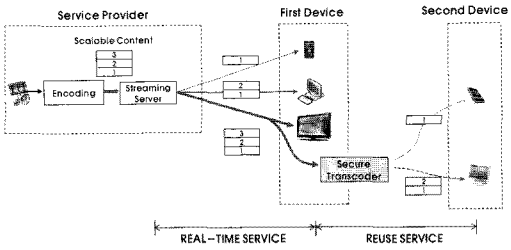
II. Related Work

The recent progress in video coding research is enabling the development of scalable video coding, which allows for almost arbitrary combinations of bitstream layers in temporal, spatial and quality dimensions. Scalable coding is to encode the video once and then lower qualities, spatial resolutions, and temporal resolutions could be obtained by simply truncating certain layers or bits from the original stream. SVC (Scalable Video Coding) is a highly attractive solution among several scalable coding schemes since it offers a variety of valuable functionalities. It has been standardized as a scalable video coding extension of the H.264/AVC standard by the Joint Video Team of the ITU-T VCEG and the ISO/IEC MPEG [1], [2]. Thus, the scalable video coding is expected to easily produce scalable content for the next generation IPTV service.

However, security remains top challenge. Conventionally, two security techniques such as CAS (Conditional Access System) and DRM (Digital Rights Management) have been used for content protection. CAS is a technique for allowing only a subscriber who is authorized to receive broadcast contents, to descramble the signals scrambled by a digital broadcast transmitter and thus to view a relevant program. Further, DRM technique is for technically protecting digital content under copyright; it is implemented such that, when a user desires to use distributed digital content which are under copyright, the digital content can be used only after obtaining license giving authorization to use the content. For interoperability between CAS and DRM technologies, an architecture which can securely convert the broadcasted content into the digital content under copyright is now being standardized [3].

However, the architecture for converting the CAS content into the DRM content requires that the content protected by CAS should be decrypted. This makes it difficult to ensure the end-to-end security of the content over the entire service range. Moreover, conventional techniques usually encrypt the entire content. This approach seems inadequate in situations where only few resources are available

(real-time networking, high-definition delivery, low memory, low power). Therefore, many functionalities of the encoding scheme may be disabled. Like this, conventional techniques for content protection have many limitations, especially in meeting new security requirements for the next generation IPTV services [4]. Some recent works explores a way of selective encryption by applying encryption to a subset of a bitstream [5]-[7].



[Fig. 1] The next generation IPTV service.

III. Security Requirements for Next Generation IPTV Services

This section defines the next generation IPTV service and then derives the security requirements to successfully deploy this service.

The next generation service is defined as a new paradigm service with 4A (Any-content, Any-device, Anytime, Anywhere) characteristic to the convergence of telecommunication and broadcasting. It seamlessly provides both the real-time broadcasting and reusable content services on diverse terminals through complex networks, as shown in [Fig. 1]. This can be efficiently achieved with the scalable video coding technology and has the following two aspects distinguished from the conventional IPTV service.

First, broadband distribution is adaptively serviced in real-time according to bandwidth variation of delivery networks and condition of user's terminals. That is, the same content can be dynamically transcoded by a combination of spatial, temporal and quality scalability. Variation in delivery networks also leads to different tradeoffs in terms of coding efficiency and error robustness, e.g. between wired and wireless networks. This media adaptation of the scalable content has an advantage of the lower server storage capacities since it is not necessary to store multiple versions of the same content. It also contributes to the simple management of the content. Consequently, this feature will surely provide practical benefits to service providers as well as consumers.

Second, a customer can directly reuse or redistribute the content even after content has been delivered to the customer's home in real-time or on-demand. From a customer's point of view, this will become a highly attractive service since the customer can reuse content on any devices, anytime and anywhere only if the customer is legally authorized. Reusable content can be easily produced by storing content in a PVR (Private Video Recorder) or a set-top box and then simply by truncating a part of the content to be suitable for a device. It thus can be redistributed to other devices through various Internet channels such as P2P file sharing, websites, messenger services.

Therefore, for successfully providing this scalable service, it is important that the following security requirements should be met in preference.

3.1 Lightweight Security

Smartphones and other mobile terminals are more widely used for multimedia service while still requiring access control and copyright protection. Their moderate resolution and computational power impose to make an effort in reducing the encryption computational complexity. In particular, real-time broadcasting services over network and public channels need to rely on access control systems to protect their content. Standard cryptographic techniques can guarantee high level of security, but lead to the cost of expensive implementation and important transmission delays. Therefore, it is necessary to perform lightweight encryption that can provide sufficient security with an important gain

in computational complexity and delays.

3.2 Perceptual Security

In traditional content protection schemes, the compressed bitstream is entirely encrypted using a standard cipher. It alters the whole bitstream syntax which may disable some codec functionalities. However, in some real-time content, it could be desirable to encourage users to buy the content. For this purpose, only a soft visual degradation is achieved, so that anyone would still understand the content but prefer to pay to access the full-quality unencrypted content. Thus, an encrypted bitstream should be compliant with the encoder; any standard decoder should be able to decode the encrypted bitstream without decryption. This can be achieved by partially encrypting specific parameters within an encoding process. Consequently, perceptual security is highly attractive, especially to service providers since it can suggest them with new business models.

3.3 Transcodable Security

After IPTV content is delivered to the customer's home in real-time or on-demand, the need for reusing the content on various devices increases more and more. This can be simply achieved by storing the content in a PVR (Private Video Recorder) or a set-top box and then securely transcoding the content according to the different capabilities of end-user's devices. However, the conventional techniques such as CAS and DRM cannot sufficiently solve this problem since the content protected by the CAS should be decrypted to be converted into the DRM content for reuse. Therefore, it is necessary that reusable content should be securely created while being stored in the PVR and then being transcoded according to the conditions of various devices without decryption.

3.4 Adjustable Security

Reusable content can be easily redistributed to other devices through various Internet channels such as P2P file sharing, websites, messenger services. If the content is obtained by an illegal user, it can cause serious problems. Hence, the encryption strength of the contents needs to be higher than that of encryption applied to real-time broadcasting. In this case, the decryption of encrypted content should not be performed during the process for converting the real-time broadcasting content into the reusable content.

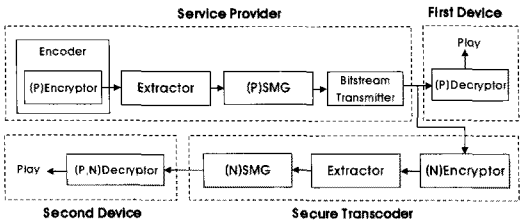
3.5 End-to-end Security

End-to-end security should be ensured over the entire range to which the IPTV service is provided. For this, a single security mechanism needs to be applied to IPTV content for integrating the real-time and reusable content services. Such mechanism should also include key management concerning tradeoff between complexity and efficiency for multiple keys.

IV. Proposed Framework

In this section, the proposed security framework is briefly reviewed and then its three main components are described in detail: parameter-based encryptor, NAL-based encryptor and security message generator.

The proposed framework is schematically shown in (Fig. 2). First, the input media content is encoded while being selectively encrypted in the service provider side. For the encryption of specific parameters, the encoding parameters (e.g., intra/inter residue signs, and a motion vector difference value) obtained during the SVC encoding process, are selectively encrypted on a per-layer basis. Then, the security message including encryption information is embedded into the encoded content, resulting in scalable secure content. This scalable secure content is broadcasted to the first device group in a form of bitstream. The security message can be identified to allow the media content to be used by the first



SMG: Security Message Generator, (P): Parameter-based, (N): NAL-based

(Fig. 2) The proposed security framework.

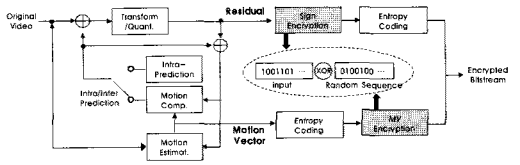
device group.

The first device can watch a real-time broadcast if it is authorized. Further, if the received content is attempted to be reused in another device, the content is first transferred to the secure transcoder while keeping encrypted. Such transcoder is located within a home, and may be operated together with a PVR or a set-top box. In the trancoder, reusable content is created by additionally performing NAL-based encryption. Here, security messages are newly generated or modified by using the existing security messages which have already been generated in the service provider. Then, the reusable secure content can be redistributed by means of various Internet channels such as Peer-to-Peer (P2P) networks or web hard drives, and be easily used by the second device.

Clearly, this reusable content is very secure since the second device should have an NAL-based decryption key as well as a parameter-based decryption key to be normally decrypted. Therefore, its security strength is usually higher than that of the content encrypted based on parameters.

4.1 Parameter-based Encryptor

The aim of parameter-based encryption is to reduce the amount of data to encrypt while preserving a sufficient level of security without altering the whole bitstream syntax. This computing savings is very desirable especially in constrained communications such as real-time networking and mobile communications with limited computational power devices. Moreover, it can produce



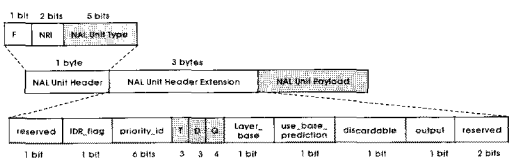
(Fig. 3) Base layer encoder with parameter-based encryptor.

quality-controllable content since the encrypted bitstream based on parameters is basically compliant with the encoder.

(Fig. 3) shows the parameter-based encryptor in the base layer. The residual sign values are encrypted before entropy coding by using exclusive-or operation with a random sequence. It can be easily generated from a pseudo random number generator with a seed value. On the other hand, motion vector difference values, which are encoded with Exp-Golomb codes, are encrypted with only bits for representing the absolute of the values during the entropy coding. However, its encryption process is similar to the method for encrypting the residual sign values. In this way, each layer can be selectively encrypted in the two domains. Consequently, this scheme can be effectively applied for satisfying both the lightweight and perceptual security requirements under the real-time broadcasting environment.

4.2 NAL-based Encryptor

NAL-based encryption is configured such that the payloads of NAL unit types 5 (IDR: Instantaneous Decoding Refresh), 1 (non-IDR), 20 (Scalable Extension), 6 (SEI: Supplement Enhancement Information), 7 (SPS: Sequence Parameter Set), and 8 (PPS: Picture Parameter Set) are selectively encrypted by using a standard cipher (DES, AES, etc.); where IDR, non-IDR, and Scalable

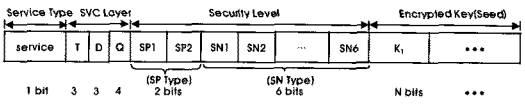


(Fig. 4) NAL unit syntax.

Extension NAL units contain information about encoded slice data while SEI, SPS, and PPS contain additional information. Such NAL unit types can be identified from NAL unit header as shown in [Fig. 4]. NAL-based encryptor is intended to additionally encrypt the received content and then directly transcode the encrypted content which might be encrypted based on parameters and NALs. As a result, it contributes to satisfying both the adjustable and transcodable security requirements, mentioned in Section III.

4.3 Security Message Generator

The bitstream of content basically includes security messages. The security message format is defined as [Fig. 5]. It is noted that at the secure transcoder, the security message is created by sharing the security message that has been generated at the service provider. This property enables one security mechanism to seamlessly be applied to both the real-time broadcasting and reuse services, resulting in end-to-end security over the entire range to which the IPTV service is provided. With the security message, encryption information can be identified to allow the media content to be used by a consumer's device.



[Fig. 5] Security message format.

Such security message is characterized in that it is generated by recording security levels and encrypted keys for each SVC layer into the payloads of one of the NAL (Network Adaptation Layer) unit types 24 to 31, which are now unspecified for the optional use. This security message can be differently set per SVC layer. Here, the security levels can be set in the message by properly combining both the encryption types based on parameters and NALs. Each encryption type is defined as shown [Table 1]. It is possible to set various security levels per SVC layer since each bit of the security levels can be independently determined with maximum 2^8 security levels.

[Table 1] Security Level

Encryption Type	Parameter
SP1	Residual Sign
SP2	MVD Value
SN1	IDR NAL Unit (5)
SN2	Non-IDR NAL Unit (1)
SN3	Scalable Extension (20)
SN4	SEI NAL Unit (6)
SN5	SPS NAL Unit (7)
SN6	PPS NAL Unit (8)

[Table 2] An Example of Security Message

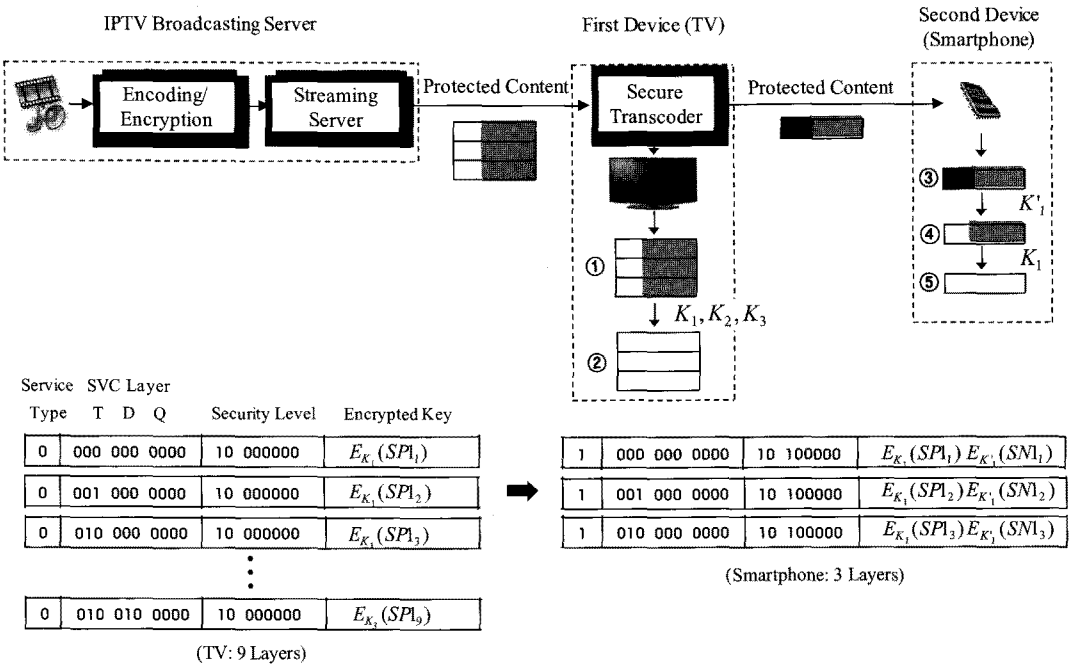
Field Name	Description
Service Type	0: real-time, 1: reuse
SVC Layer	001 000 0001: (T,D,Q) = (1,0,1) layer
Security Level	10: SP1 encryption 100000: SN1 encryption
Encrypted Key	1 encrypted seed: SP1 1 encrypted key: SN1

Similarly, encryption seeds or keys are encrypted by using a standard cipher and then the encypted values are set into the message. The number of the encrypted keys equals to the number of bits to be set in the security level. When the bitstream of content is generated, the NAL unit with the security message is located at the beginning of the bitstream. This allows a device to interpret the NAL unit prior to other NAL units. [Table 2] shows an example of this security message.

V. Experiments

In this section, the validity of the proposed framework is evaluated through the experiments using the test sequence "ICE" to the JSVM9.19 [8], [9].

A service model is here assumed as shown in

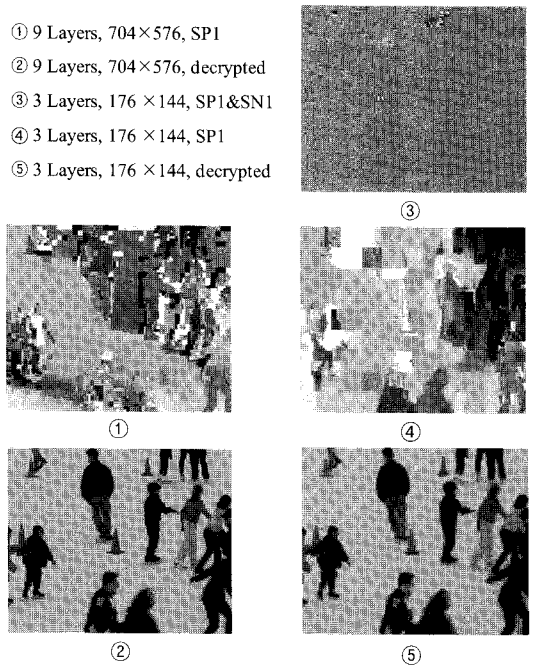


(Fig. 6) An service scenario and security messages.

(Fig. 6) where the scalable content with 3 spatial scalabilities (4CIF, CIF, QCIF), 3 temporal scalabilities and 1 quality scalability is encoded. The content is simultaneously encrypted with SPI encryption type defined in (Table 1), and the seed in each layer is generated differently, resulting in 9 different seeds for overall 9 layers. However, for simple key management, such seeds are encrypted with only three keys (K_1, K_2, K_3) according to the spatial scalabilities. As shown in (Fig. 6), the security messages are generated and then distributed while being embedded in the media bitstream. It is noted that at the secure transcoder, the base layer of this content is additionally encrypted with SNI encryption type using the key K'_1. Then, only the base layer of the content is extracted in the secure transcoder for the reuse in the second device. Here, the security messages are also extracted and then modified as shown in (Fig. 6).

In this scenario, the broadcasting server first creates the protected content and then distributes the content to a consumer. If the consumer has paid for the content, the broadcasting server delivers the three authorized keys to the first device. Therefore, the device can successfully decrypt and play the encrypted content after easily obtaining the overall 9 different seeds only if it has the three authorized keys corresponding to the three spatial layers.

To evaluate the practicality of the proposed framework, experiments were performed at the five points numbered in (Fig. 6). (Fig. 7) shows the experimental results. That is, the second and fifth results demonstrate the images decrypted and decoded normally in TV and smartphone, respectively. The first result is the image decoded in TV without the keys (K_1, K_2, K_3), which is easily obtained owing to format compliant property of parameter-based encryption. Similarly, the fourth result is the image decoded in smartphone without the key K_1. When looking carefully at these two images, it is noticed that the scene is perceptually intelligible although they are quite



[Fig. 7] Decoded results at the five points of [Fig. 6].

noisy. On the other hand, the third image reveals no meaningful information. Since the original bitstream of this image is altered by the NAL-based encryptor, it cannot be directly decoded by using any standard decoder. Therefore, only the bits compliant with the SVC codec are decoded without the keys (K_1, K'_1) while replacing the noncompliant bits with a default value. It is suggested that the encryption strength of the content can be controlled to an almost unintelligible level by selectively using the encryption types.

The experiments were additionally performed using the standard video sequence "HARBOUR" which was encoded with the same configuration used in the "ICE" sequence. In real-world service environments, decryption is more critical rather than encryption because the encryption overhead can be sufficiently overcome by high-performance computing systems. Thus, the computational overhead during the decrypting process was measured instead of the encryption process.

[Table 3] Time Overhead at Each Decryption Point

Sequence	①→②	③→④	④→⑤
ICE	0.28%	0.17%	0.27%
HARBOUR	0.35%	0.45%	0.25%

[Table 3] shows the time overhead at three decryption processes in [Fig. 6]. This is a time ratio between decryption and decoding. Here, the computer used is based on a 2.67 GHz Intel Processor and with 2.75 GB of RAM. As a result, the decryption overhead can be said to be negligibly small compared to the decoding time.

VI. Conclusion

The next generation IPTV service has notable features such that it can seamlessly provide the real-time broadcasting service and the reusable service. In this paper, a new scalable security framework suitable for this service is proposed. The proposed framework has many advantages in that it can potentially create various business models by using its lightweight, perceptual, transcodable, and adjustable encryption properties compared to conventional content protection techniques. Then, experiments under the practical service scenario are performed. The test results show the suitability of the proposed framework.

In the future, the implementation supporting more service models will be attempted. Moreover, the techniques of self-protecting, watermarking and fingerprinting will be studied for more securely protecting and exactly tracing the reusable content.

References

[1] ISO/IEC JTC 1/SC 29/WG 11 and ITU-T SG16 Q.6: Scalable Video Coding - Joint Draft 11, Doc.JVT-X201 July, 2007.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," IEEE Trans. on Circuit and Systems for Video Technology 17(9), pp. 1103-1120,

Sep. 2007.

[3] ITU-T IPTV-GSI, http://www.itu.in/ITU-T/iptv

[4] Seungmin Lee, Jaehoon Nah, and Dongil Seo, "Security requirements for scalable IPTV services," The Proc. of the 2009 International Conf. on Security and Management, Vol. 1, pp. 99-102, July. 2009.

[5] Y. G. Won, T. M. Bae, and Y. M. Ro, "Scalable protection and access control in full scalable video coding", IWDW 2006, LNCS 4283, pp. 407-421, 2006.

[6] S.-W. Park, S.-U. Shin, "Efficient selective encryption scheme for the H.264/Scalable Video Coding (SVC)", 2008 Fourth International Conf. on Networked Computing and Advanced Information Management, pp. 371-376, Sep. 2008.

[7] A. Massoudi, F. Lefebvre, D. D. Vleeschouwer, B. Macq, and J.-J. Quisquater, "Overview on selective encryption of image and video: challenges and perspectives," EURASIP Journal on Information Security, Vol. 2008, Article ID 179290, 18 pages, 2008.

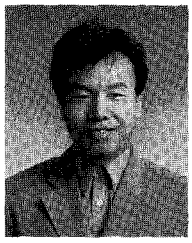
[8] Test Media, http://media.xiph.org/video/derf/

[9] ISO/IEC JTC 1/SC 29/WG 11 N8750: Joint Scalable Video Model (JSVM), Jan. 2010.

〈著者紹介〉

이 승 민 (Seungmin Lee) 정회원
1985년 2월: 고려대학교 산업공학과 졸업
1997년 2월: KAIST 산업공학과 석사
2010년 8월: KAIST 산업및시스템공학과 박사
1997년~2000년: 데이콤 종합연구소 주임연구원
2001년~현재: 한국전자통신연구원 선임연구원
〈관심분야〉 정보보호, Anomaly Detection, 데이터마이닝

나 재 훈 (Jae Hoon Nah) 정회원
1985년: 중앙대학교 컴퓨터공학과 졸업
1987년: 중앙대학교 컴퓨터공학과 석사
2005년: 한국외국어대학교 전자정보공학과 박사
1987년~현재: 한국전자통신연구원 책임연구원
〈관심분야〉 IPv6/MIPv6 보안, P2P 보안, IPTV 보안



서 동 일 (Dong-Il Seo) 종신회원
1989년 2월: 경북대학교 전자공학과 졸업
1994년 2월: 포항공과대학교 정보통신공학과 석사
2004년 2월: 충북대학교 전자계산학과 박사
1994년~현재: 한국전자통신연구원 팀장 (책임연구원)
2010년~현재: 충남대학교 컴퓨터공학과 겸임교수
2002년~현재: ASTAP-Forum 정보보호EG, ITAU워킹그룹 의장
1994년~현재: TTA 정보보호/네트워크 표준화 (현 TC5 부의장)
〈관심분야〉 Network, 미래 인터넷, 정보보호 (네트워크보안, 해킹 등)