A Hierarchical Encryption and Key Management Scheme for Layered Access Control on H.264/SVC Bitstream in the Internet of Things

Cheng Xu, Wei Ren, *Member, IEEE*, Linchen Yu, Tianqing Zhu, *Member, IEEE*, and Kim-Kwang Raymond Choo, *Senior Member, IEEE*

***Abstract*—Terminals with diverse technological specifications, heterogeneous network environment, and personalized user requirements raise new challenges to streaming media services. Solutions such as the newly standardized H.264/SVC (scalable video coding; designed to compress original video bitstream into a multilayer video stream according to requirements) have been proposed. With the pervasive application of SVC in applications, such as video on demand, video conferencing, and video surveillance in the Internet of Things (IoT), there has been increased scrutiny on security of H.264/SVC. In this article, we propose a bitstream-oriented layered encryption scheme for SVC bitstream. According to the multilayer bit code structure of SVC, the bitstream is separated and encrypted, respectively, by rearranging the network abstraction layer (NAL) unit of SVC bitstream. This provides hierarchical protection for the multilayer characteristic of SVC. In order to provide sufficient security, as well as achieving improved computational efficiency, we use different cryptographic algorithms for the base layer and enhancement layers according to its requirements. The base layer adopts off-the-shelf high-security encryption algorithms, such as block cipher, to ensure security. Each enhancement layer is encrypted with a different key through the stream cipher with low computational complexity, providing layered control of the video. Furthermore, we propose a hierarchical key management scheme to implement layered access control according to the principle of hierarchical deterministic wallet (H-D wallet). Our scheme can be applied to the user-level distinction in video on demand and video surveillance systems in IoT. The analysis and experiments indicate that the proposed scheme achieves a high-security level, yet incurs reasonably low compression cost and computational complexity.**

Manuscript received November 25, 2019; revised February 13, 2020; accepted May 20, 2020. Date of publication May 26, 2020; date of current version September 15, 2020. This work was supported in part by NSFC under Grant 61972366 and Grant 61502439, in part by the Major Scientific and Technological Special Project of Guizhou Province under Grant 20183001, and in part by the Open Funding of Guizhou Provincial Key Laboratory of Public Big Data under Grant 2018BDKFJJ009 and Grant 2017BDKFJJ006. *(Corresponding author: Wei Ren.)*

Cheng Xu is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China.

Wei Ren is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China, and also with the Guizhou Provincial Key Laboratory of Public Big Data, Guizhou University, Guizhou 550025, China (e-mail: weirencs@cug.edu.cn).

Linchen Yu is with the School of Cyber Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China.

Tianqing Zhu is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China, and also with the School of Software, University of Technology Sydney, Sydney, NSW 2007, Australia.

Kim-Kwang Raymond Choo is with the Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249 USA.

Digital Object Identifier 10.1109/JIOT.2020.2997725

***Index Terms*—H.264/SVC, hierarchical key management, Internet of Things (IoT), layered access control, layered encryption.**

I. Introduction

ACCORDING to the Cisco visual networking index (VNI) report, the distribution and transmission of the Web content have increased significantly, among which video traffic accounted for 82% [1]. In our interconnected and media-rich society, the broad range of terminals with varying technological specifications, heterogeneous network environment, and personalized user requirements complicate streaming media services. The scalable video coding (SVC) extension of the H.264-advanced video coding standard (H.264/AVC), for example, is one of the several initiatives designed to cater to the new networking environment [2].

H.264/SVC is a standard issued by the joint video coding group (JVT) in 2007 [3], which inherits the encoding mode of H.264/AVC with high coding efficiency [4]. H.264/SVC supports spatial, temporal, and quality-scalable coded videos, and can compress the original bitstream into a multilayer video stream according to requirements. The SVC bitstream consists of a base layer and several enhancement layers. The lowest video quality can be obtained by decoding the base layer, and enhancement layers are added on the base layer to obtain a higher quality video. Therefore, SVC can deal with the heterogeneous network environment and provide diversified services according to user needs [5], [6].

With the extensive application of SVC in video on demand, video conferencing, video surveillance [7]–[9], and other scenarios, a content protection scheme for SVC bitstream is needed to solve the security problem of a scalable video, such as copyright protection of video and encryption for video conferencing. In addition to the protection of the video content, a corresponding protection mechanism for layered access control of SVC video should be developed. For example, surveillance systems become an important part of the Internet-of-Things (IoT) era [10]. As far as surveillance video of IoT, low-privilege users are often restricted to view merely the outline of the video, while high-privilege users can access the high-definition video sequence. Moreover, in the video-on-demand system, ordinary users and paying users need

to be treated separately. Ordinary users can access only low-quality videos while paying users can access all the bitstreams of video (base layer and enhancement layers) [11]. The layered access control for SVC can be applied in these scenarios.

In the existing encryption scheme for SVC bitstream, some of them select some pivotal coefficients of the encoding process to encrypt, which avoids the compression overhead but ignores the security [12]–[15]. Others use the cryptographic algorithm to encrypt bitstream without considering the different significance between the base layer and enhancement layers. Moreover, the existing schemes only focus on computational complexity and security, which do not provide layered access control of SVC [16]–[19]. This article proposes the hierarchical encryption and key management scheme for layered access control on H.264/SVC bitstream. Compared with the existing schemes, the proposed scheme utilizes the hierarchical structure of SVC to encrypt the base layer and the enhancement layers by different algorithms, respectively. Furthermore, a hierarchical key management scheme is proposed to generate and manage the keys used in the adopted encryption algorithms. Our scheme implements the layered access control of SVC, which can be applied to the user-level distinction in video-on-demand and video surveillance systems in IoT. The main contributions of this article are as follows.

1) This article splits and recombines the network abstraction layer (NAL) unit of SVC bitstream to implement the separation of layers. According to the differences in the base layer and the enhancement layers, different encryption algorithms are used to ensure the security as well as improve the encryption efficiency.
2) The base layer adopts the block cipher algorithm with high security to ensure the security of the video content. Each enhancement layer is encrypted with a different key through stream cipher with low computational complexity, providing quality-level control of the video. This article implements the hierarchical protection of SVC without sacrificing the scalability.
3) A hierarchical key management scheme is proposed according to the principle of a hierarchical deterministic wallet (H-D wallet). High-level users can derive lower level keys while the reverse process is not feasible. It not only reduces the complexity of key management but also ensures the security of the keys used in the layered encryption scheme.
4) Our scheme provides the layered access control of SVC. It guarantees that unauthorized users cannot watch the video, and the quality of video content accessed by users depends on their privilege level, which is applicable to video surveillance systems of IoT.

The remainder of this article is organized as follows. The related literature is briefly reviewed in Section II. The SVC bitstream structure and design goals are introduced in Section III. In Sections IV and V, we describe our proposed encryption scheme and present its experimental results, respectively. Finally, conclusions are drawn in Section VI.

II. Related Work

There are a number of security challenges underpinning the streaming of media data, and one ongoing research agenda is to design efficient and secure video encryption algorithms such as those based on earlier video standards. The existing video encryption algorithms can be broadly categorized into those based on compression-integrated encryption and those based on bitstream-oriented encryption.

Compressed-integrated encryption schemes integrate the encryption scheme into the process of video compression, which selects some pivotal coefficients to encrypt. When the video encoding is completed, the video data have been encrypted. Asghar *et al.* [20] encrypted the codewords of the signs of T1s, EG0 suffix, and the bin-strings of UEG0 suffix with signs of TC levels in the entropy coding process. The schemes proposed by Won *et al.* [12] encrypt the signs of texture, intramode, and motion vectors during SVC encoding. Similarly, the schemes proposed by Park and Shin [13], [14] and Kim *et al.* [15] encrypt the signs of motion vectors, the signs of intraprediction modes, and the signs of residual coefficients. Based on the research of interlayer residual prediction, Liu *et al.* [21] found that the decoding of macroblocks of the enhancement layer must use the residual signal of the base layer. Hence, only the nonzero quantization residual coefficient of the base layer is encrypted in their scheme, which improves the encryption efficiency. On the contrary, Algin and Tunali [22] encrypted both signs and DC coefficients to improve security but decreased the compression rate by 15%.

One important disadvantage of compressed-integrated encryption is its lack of flexibility because encryption adjustments require the recoding of the video itself [2]. The bitstream-oriented encryption directly encrypts the encoded video bitstream. So modifying the encryption algorithm does not require recompression. The encryption and decryption steps are also more simplified. H.264/SVC bitstream consists of many NAL units, which can be encrypted according to the bitstream structure. In the scheme proposed by Li *et al.* [16], the NAL units of instantaneous decoding refresh (IDR) picture, picture parameter set (PPS), or sequence parameter set (SPS) are ciphered by the ex-leak extraction (LEX) algorithm. In the schemes proposed by Hellwagner *et al.* [17] and Stütz and Uhl [18], they mapped the NAL unit (NALU) types of an encrypted SVC bitstream to unspecified NAL unit types (NUTs), i.e., types 24–27. But a third-party decoder ignores the unspecified NUTs, it can parse the encrypted bitstream smoothly. Wei *et al.* [19] created new NALUs to replace original video coding layer (VCL) NALUs by encrypting VCL NALUs into either supplement enhancement information (SEI) NALU. At the same time, each new NALU includes an SVC NAL header to indicate the scalability information, which satisfies the SVC specification.

Considering that none of the above methods provides hierarchical protection for the multilayer characteristic of SVC, a bitstream-oriented layered encryption scheme is necessary.

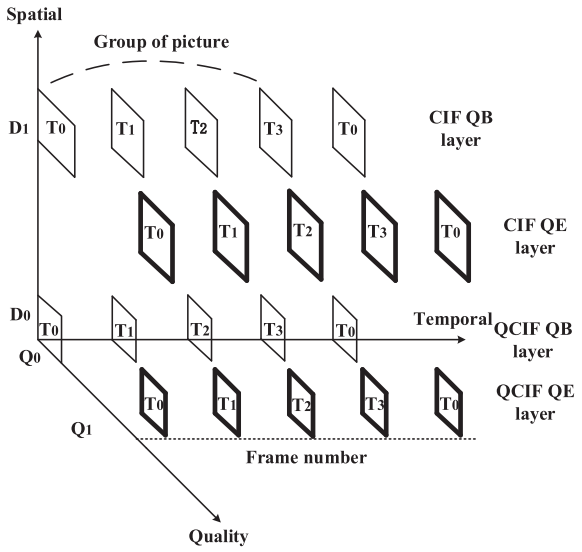


Fig. 1. Composite structure of SVC.

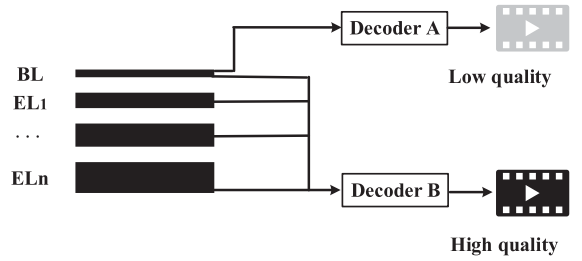


Fig. 2. SVC bitstream extraction rules.

TABLE I
NAL TYPE

nal_unit_type	Content of NAL unit and RBSP syntax structure	C
0	Unspecified	
1	Coded slice of a non-IDR picture	2,3,4
2	Coded slice data partition A	2
3	Coded slice data partition B	3
4	Coded slice data partition C	4
5	Coded slice of an IDR picture	2,3
6	Supplemental enchancement information (SEI)	5
7	Sequence parameter set	0
8	Piceture parameter set	1
9	Access unit delimiter	6
10	End of sequence	7
11	End of stream	8
12	Filler data	9
13...23	Reserved	
24...31	Unspecified	

TABLE II
NOTATIONS

Notations	Meaning
SVC	Scalable Video Coding
AVC	Advanced Video Coding
VCL	Video Coding Layer
NAL	Network Abstract Layer
NALU	Network Abstract Layer Unit
XOR	Exclusive OR
H-D wallet	Hierarchical Deterministic Wallet
BL	Baes Layer
EL	Enhancement Layer
DEC	Decipher
ECC	Elliptic Curve Encryption
AES	Advanced Encryption Standard
QCIF	Quarter Common Intermediate Format(176x144)
CIF	Common Intermediate Format(352x288)
QP	Quantitative Parameters

III. PRELIMINARIES

A. SVC Bitstream Structure

The H.264/SVC bitstream consists of a base layer and several enhancement layers. The base layer is compatible with the H.264/AVC standard, and the enhancement layers support temporal, spatial, and quality scalability to increasing the frame rate, resolution, and quality of the video. Moreover, H.264/SVC provides a relatively complete layered mechanism that can flexibly implement any combination of the three extensions. The classification is based on the spatial layer, and each spatial layer can contain multiple quality layers. Fig. 1 shows a composite structure with four temporal layers, two space layers, and two quality layers, where T_0 represents the temporal base frame. One base frame and all the enhancement frames between it and the next base frame constitute a Group of Picture (GOP).

The multilayer SVC bitstream allows users to extract videos of different qualities according to their requirements. As shown in Fig. 2, decoder *A* only obtains low-quality video sequences, while decoder *B* decodes more enhancement layers to improve the quality of the video. The enhancement layers take effect under the premise of decoding the base layer.

H.264/SVC and H.264/AVC have similar bitstream structure, which can be divided into two layers according to functions: 1) VCL and 2) NAL. VCL stores video content and VCL data are encapsulated in NAL units before transmission. Each NAL unit includes a raw byte sequence payload (RBSP) and a set of NAL headers information. NAL units are separated by separate bytes. The NAL header occupies 1 B (after adding the svc extension and occupying 4 B), contains the NAL unit type information (nal_unit_type) and some other information, the NAL type is shown in Table I.

14 to 20 are used as SVC tags, and the extended header information mainly includes D (spatial hierarchy number), Q (quality hierarchy number), T (temporal hierarchy number), and other information. The NAL units of the same frame in the SVC have the same time level. The space priority is ranked first and the quality level is ranked second. Then, the adjacent frames are arranged from small to large according to the time level.

B. Notations

To better understand this article, Table II shows the key abbreviations in the text.

C. Design Goals

The design goals of our scheme are as follows.

1) Encrypt all base and enhancement layers for the robustness of video security.

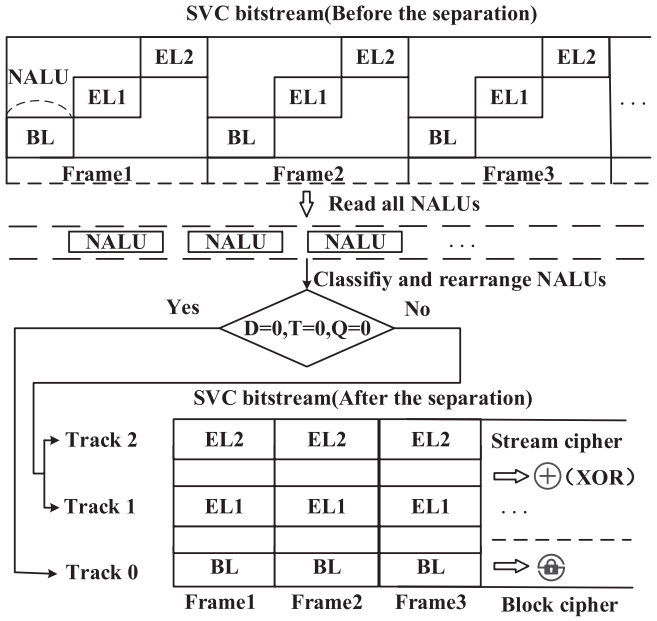


Fig. 3. Layered encryption.

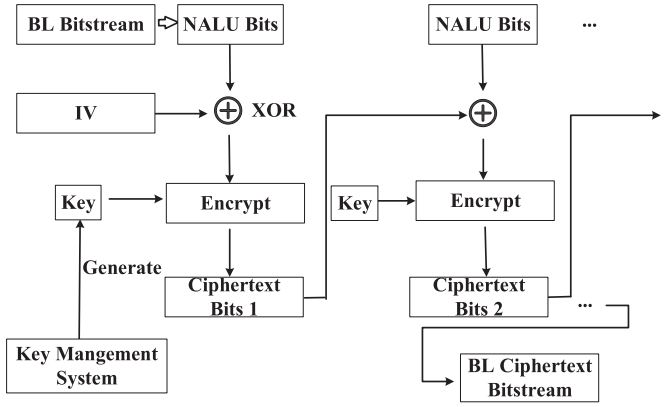


Fig. 4. Base layer encryption.

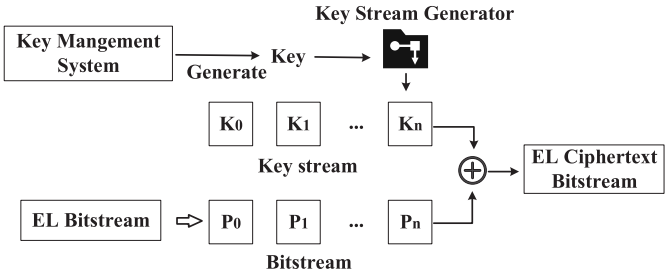


Fig. 5. Enhancement layer encryption.

2) Acceptable compatibility. Even if the encryption algorithm is replaced, it does not need to recompress the video.
3) In order to ensure sufficient security and improve the computational efficiency, the base layer and the enhancement layers adopt different encryption algorithms according to its significance.
4) Hierarchical protection is provided. The base layer is protected by the high-security encryption algorithm. Without the key of the base layer, the SVC bitstream cannot be decoded. The enhancement layers control the quality of the video. Without the keys of the enhancement layers, users only get the video code rate of a low-level layer after decoding.
5) Provide a key management scheme to ensure that keys used in the layered encryption scheme are sufficiently secure and provide layered access control.

IV. Proposed Scheme

Based on the layered characteristic of SVC bitstream, this article proposes a layered encryption scheme for SVC bitstream to provide hierarchical protection. The base layer and the enhancement layers adopt different encryption algorithms according to their significance. Then, we propose hierarchical key management to generate and manage all keys based on the principle of the H-D wallet. Finally, a layered access control algorithm is proposed.

A. Layered Encryption Scheme

As shown in Fig. 3, all NAL units of SVC are separated and reorganized. In accordance with the DQT classification, NALUs with the same DTQ level are recombined according to the original relative order to implement layers separation. At the same time, when detecting $D = 0$, $T = 0$, $Q = 0$, and the base layer code stream, the whole NALU is encrypted by the block encryption algorithm. Otherwise, the separated enhancement layer NALUs is encrypted by stream cipher with different keys.

Since the base layer is the most important in SVC bitstream, the block encryption algorithm with high security and efficiency is selected for encryption to ensure the security of the base layer. As shown in Fig. 4, the block encryption operates in the CBC mode. After grouping the base layer bitstream, the first block is XOR with the randomly generated initialization vector IV, then encrypted with the key generated by the key management system. The generated ciphertext block is used as the initialization vector of the second block. Until all NALU units are fully encrypted, we get the ciphertext bitstream of the base layer. The encryption process is defined as

$$M_0 = E(C_0 \oplus iv, K) \tag{1}$$

$$M_i = E(C_i \oplus M_{i-1}, K)(i = 1, 2, \ldots) \tag{2}$$

where C_i is the base layer bitstream, K is the encryption key, iv is the initialization vector, and M_i is the ciphertext bitstream.

In terms of enhancement layers, the importance of the enhancement layer is lower than that of the base layer. In order to reduce the computational cost, this article utilizes a lightweight stream cipher encryption scheme to encrypt each enhancement layer with different keys. Fig. 5 shows the process of encrypting an enhancement layer. The key generated by the key management system generates the keystream through a keystream generator. The same key generates the same keystream. Then, the code stream of the enhancement layer is XOR with the stream key. So we can quickly encrypt the bitstream of the enhancement layer to improve the overall

Algorithm 1: Layered Encryption

Input: the SVC bitstream S, layer number n, the base layer BL=null, the enhancement layer EL_i=null($i = 1, ..., n$), encryption key $K_i (i = 0, ..., n)$

Output: the encrypted bitstream of each layer BL, $EL_i (i = 1, ..., n)$

```
Read the N_i(NALUs, i = 0, ..., m) of the bitstream;
while N_i do
    if N_i.DTQ = 000 then
        BL.append(N_i)
    else
        Switch(N_i.DTQ)
        Case V_1: EL_1.append(N_i); break;
        ...
        Case V_n: EL_n.append(N_i); break;
    end
end
for (i=0;i ≤ n;i++) do
    if i = 0 then
        BL = E_AES(K_0, BL)
    else
        EL_i = E(K_i ⊕ EL_i)
    end
end
```

encrypt efficiency. The encryption process is defined as

$$M_i = K_i \oplus P_i \tag{3}$$

where K_i is the stream key, P_i is the bitstream of the enhancement layer, and M_i is the ciphertext bits.

The specific layered encryption schemes are shown in Algorithm 1, where V_i–V_n means different sequences of DQT. For the block encryption, we select the AES algorithm as an example. According to the DQT value of N_i, all N_i are classified and arranged to different tracks. Then, we obtain the bitstream of each layer, BL–EL$_n$. For BL, we use AES encryption algorithms. As far as EL$_1$–EL$_n$, we adopt a stream cipher algorithm with different keys. Finally, we obtain the encrypted bitstream of each layer.

B. SVC Bitstream Decryption

Decryption is the reverse process of the encryption scheme, which decrypts the encrypted bitstream according to the keys you hold. As shown in Fig. 6, when the key of the base layer is held, you can only get the base layer of the video. The decoded video quality is increased by one level with each additional key of the enhancement layer. But the Ith layer key only takes effect when the $I-1$th layer key is held. Therefore, you can only decode the SVC bitstream to the $I-1$th enhancement layer at most when you lose K_i. If the base layer key is missing, the entire video stream cannot be decoded properly.

C. Hierarchical Key Management

In our encryption scheme, an N-layer SVC bitstream requires N encryption keys, namely, K_0–K_{n-1}. Based on the layered access control of this scheme, the user at the Ith layer with K_0–K_{i-1} cannot view the entire quality of the video sequence. Because there is no higher layer key K_i–K_{n-1}. Since each layer is assigned a different key, high-level users need to consume a lot of resources to store and manage all keys, and the loss of a key can also cause the imbalance of the layered access control. In order to mitigate this issue and prevent frequent key backup, based on the principle of an H-D Wallet proposed in the BIP32 standard [23], this article proposes a hierarchical key management scheme to generate, manage, and distribute the encryption keys of each layer, which reduces the number of keys to be saved while ensuring the security performance.

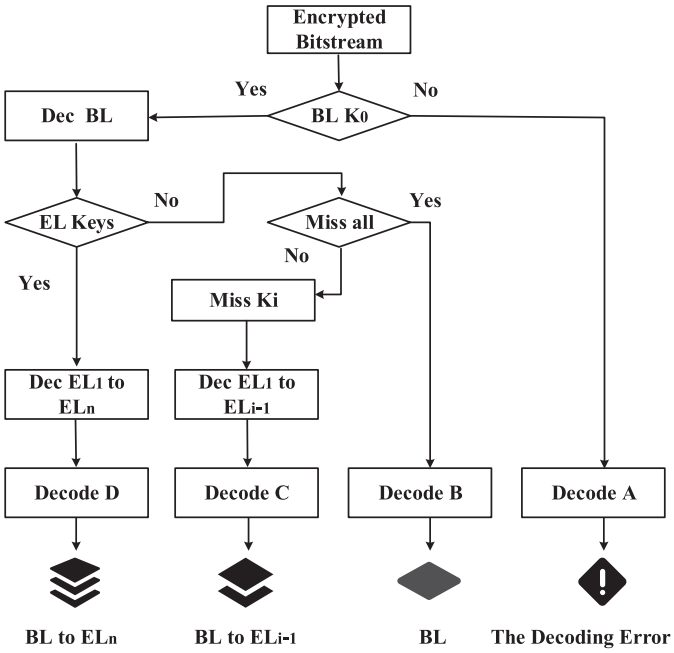


Fig. 6. Decryption.

H-D Wallet allows the parent keys to extend child keys and child keys extend grandchild keys and so on, thus performing a tree-like structure [24]. As shown in Fig. 7, the mnemonic words are used to generate the key seed through the key stretching function and 2048 rounds of the HMAC function. Then, the key seed as entropy generates the main private key and the main chain code through the Hmac-sha512 function. The main private key can generate the main public key by elliptic curve encryption (ECC). The private key and chain code can be encoded into the extended private key by Base58Check, and the extended private key can also be restored to the private key and chain code. In order to minimize the number of user-managed keys, the private key is used as the encryption key and the extended private key as the distribution key. Therefore, the highest layer key of SVC is generated by the initial seed, in which the main private key is used as the encryption key. Then, the highest layer distribution key is generated with the main private key and the main chain code by the following equation:

$$(K(\text{pri})_{n-1}, K(\text{chain})_{n-1}) = H(\text{seed}) \tag{4}$$

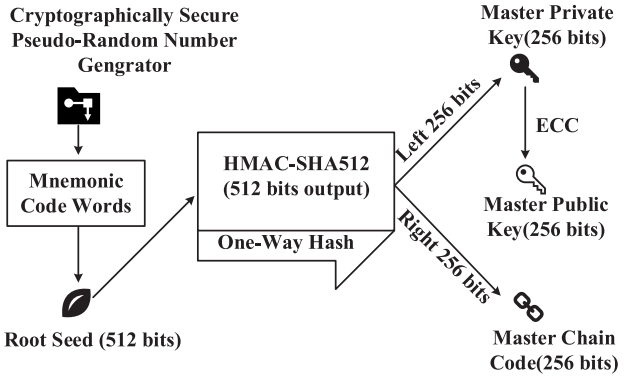


Fig. 7. Master key generation.

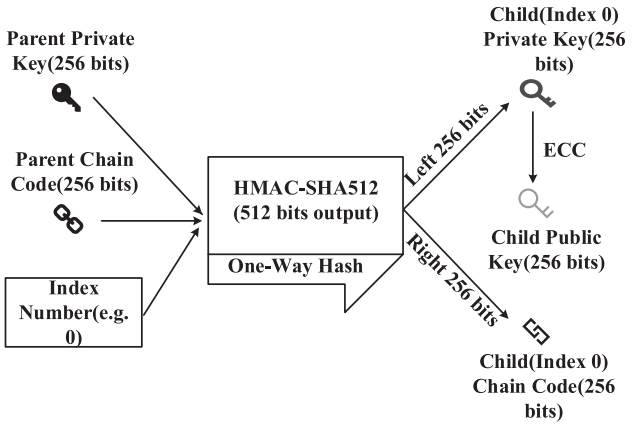


Fig. 8. Derived subkey.

Algorithm 2: Layered Access Control

Input: user key K_i, layer number n, Stored video sequences that need to be decoded (VS) = null
Output: user level UL, The final video sequence $Video$

```
if no K_i then
    UL = null, Video = null
    output cannot access the video;
else
    if i=0 then
        D_AES(BL, K_i);
        UL=0,Video=Decode(BL);
    else
        K(pri)_i = RE(K_i).private;
        K(chain)_i = RE(K_i).chaincode;
        for (;i ≥ 1;i- -) do
            D(EL_i ⊕ K(pri)_i);
            VS.append(EL_i);
            (K(pri)_{i-1}, K(chain)_{i-1}) =
              H(K(pri)_i, K(chain)_i, n − i);
            if i=1 then
                K(pub) = K(pri)_1 * G;
                D_AES(BL, K(pub));
                VS.append(BL);
            end
        end
        UL=i,Video = Decode(VS)
    end
end
```

$$K_{n-1} = BC(K(\text{pri})_{n-1}, K(\text{chain})_{n-1}). \tag{5}$$

Assume that the SVC video has N layers, where $H()$ is denoted as a cryptographic hash function. The seed is the initial seed generated by mnemonic words. $K(\text{pri})_{n-1}$ is the Nth layer private key and $K(\text{chain})_{n-1}$ is the Nth layer chain code. K_{n-1} is the Nth layer extended private key.

As shown in Fig. 8, the parent chain code is used as the entropy of generating a child key to improve the unpredictability of the child key. Then, the parent key, parent chain code, and child key number can derive child private key and child chain code through Hmac-sha512 in the method of hardened derivation. So the encryption key of the Ith layer can be derived from the following equation:

$$(K(\text{pri})_i, K(\text{chain})_i) = RE(K_i) \tag{6}$$

$$(K(\text{pri})_{i-1}, K(\text{chain})_{i-1}) = H(K(\text{pri})_i, K(\text{chain})_i, n-i) \tag{7}$$

where K_i is the extended private key of the $I+1$th layer, and $RE()$ is the operation of restoring the extended key to the private key and chain code. $n-i$ is the key number of the Ith layer, and $K(\text{pri})_{i-1}$ is the encryption key of the Ith layer. Following this logic of deduction, the key of the $I-2$th layer can be derived from $K(\text{pri})_{i-1}$ and $K(\text{chain})_{i-1}$ according to (7) until the first enhancement layer (namely, the second layer). Due to the one-way characteristic of the hash function, only high-level users can derive lower level keys while the reverse process is not feasible. Then, the public key generated by the private key of the first enhancement layer is used as the base-layer key. The key of the base layer is derived on the following equation:

$$K(pub)_1 = K(\text{pri}) * G \tag{8}$$

$$K_0 = K(pub)_1 \tag{9}$$

where G is the generation point of ECC, K_0 is the key of the base layer, and $K(pub)_1$ means the public key of the first enhancement layer. Due to the one-way of ECC, the public key cannot derive the private key. Therefore, the key of the first enhancement layer cannot be obtained with only the key of the base layer. This key management scheme not only guarantees the security of the keys used in the layered encryption scheme but also greatly reduces the complexity of key management.

D. Layered Access Control

Based on the hierarchical encryption and key management scheme, we propose a layered access control algorithm of SVC. For the block encryption, we select the AES algorithm as an example. The specific layered access control algorithm is shown in Algorithm 2.

If there is no K_i, the user cannot access the video. $i=0$ means that K_i is the key of BL. After decrypting with the AES algorithm, the user can only decode the BL. If $i>0$, K_i is restored to the private key $K(\text{pri})_i$ and chain code $K(\text{chain})_i$ by $RE(K_i)$. $K(\text{pri})_i$ is used to decrypt EL_i. Then,

TABLE III
SVC VIDEO SEQUENCES

	frame	frame rate/fps	layer	resolution	QP
Bus	75	15	B	QCIF	36
			E/sptial	CIF	36
Fate	122	30	BL	CIF	36
			E1/quality	CIF	28
			E2/quality	CIF	24
City	300	30	BL	QCIF	36
			E1/quality	QCIF	28
			E2/sptial	CIF	36
			E3/quality	CIF	28

$K(\text{pri})_i$, $K(\text{chain})_i$, and $n - i$ are used to generate $K(\text{pri})_{i-1}$ and $K(\text{chain})_{i-1}$ by a hash function $H()$. Following this logic of deduction, all low-level keys can be derived. So EL_i–EL_1 are decrypted. When $i = 1$, $K(\text{pri})_1$ can derive the key of BL by ECC. Then, the user can decode BL–EL_i.

When a user with i-level privilege accesses the video content, only the extended private key of layer I, namely, K_{i-1} is distributed. Then, the user can derive all lower level keys for decryption with K_{i-1}, but cannot obtain the higher level keys. As far as the lowest level user, we just distribute the public key of the first enhancement layer. Users of any level only need to manage one key, and users without any key cannot access the video. This key management scheme conforms to the layered access control of SVC, and the quality of the video content accessed by users depends on their keys (namely, authority or payment).

V. EXPERIMENTS AND ANALYSIS

In this section, we will evaluate and discuss our proposed scheme in terms of security analysis, performance analysis, use case, and comparison with other SVC encryption schemes.

We have implemented the proposed method in JSVM 9.19 [25]. For block encryption, we select the AES algorithm to encrypt the base layer. As listed in Table III, the video sequences include Bus (75 frames), Fate (122 frames), and City (300 frames), where B/E means the base layer B and the enhancement layer E. Fate contains a base layer and two quality enhancement layers. The resolution is CIF and the frame rate is 30 frames/s. The bus contains one base layer, one spatial enhancement layer. The resolution of the base layer is QCIF, the resolution of the enhancement layer is CIF, and the frame rate is 30 frames/s. The city contains one base layer, one spatial enhancement layer, and two quality enhancement layers. The resolution of the base layer is QCIF and that of the spatial enhancement layer is CIF.

A. Security Analysis

All keys used in layered encryption are generated by the hierarchical key management. The seed generated by 2048 rounds of hash increases the unpredictability of the highest layer key. Then, the parent chain code as entropy improves the security of child keys. The one-way characteristic of the hash function ensures that the low-level key cannot recover the high-level key.

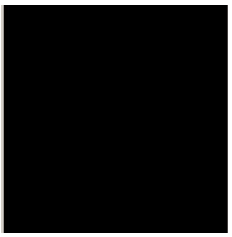


Fig. 9. Video sequence without decrypting BL.

(a)

(b)

Fig. 10. Video sequence after decrypting (a) BL or (b) BL and EL.

For the base layer, we use the AES-CBC encryption algorithm with a 256-b key. The attacker's maximum attempt space is 2^{256}. Presently, the computing power of the computer cannot crack such a long encryption key, which means the base layer has extremely high security. Therefore, as long as the key of the base layer is missing, the users cannot access the entire SVC stream. Taking the bus video sequence as an example. As shown in Fig. 9, the bus sequence is decoded without decrypting the base layer, and each frame of the SVC bitstream cannot be decoded normally. Thus, we obtain a damaged video sequence. Since the encryption of the base layer can guarantee the access control of the SVC video, for the enhancement layer, we use stream cipher with low complexity of encryption and decryption to implement the level control of video. Each layer uses different stream keys to enhance the robustness of the entire video bitstream. Moreover, after decryption with the correct keys, the video bitstream is still scalable. As shown in Fig. 10(a) and (b), Fig. 10(a) shows the video image which is decoded after only decrypting the base layer, and the resolution of the video is QCIF while Fig. 10(b) shows the video

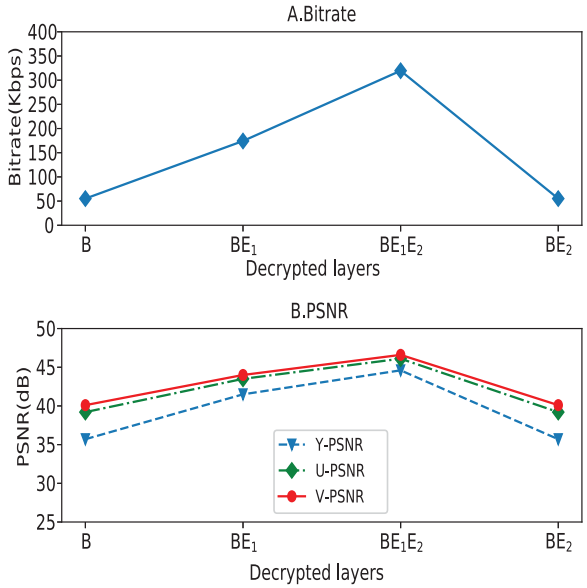


Fig. 11. PNSR and bitrate after decrypting different layers.

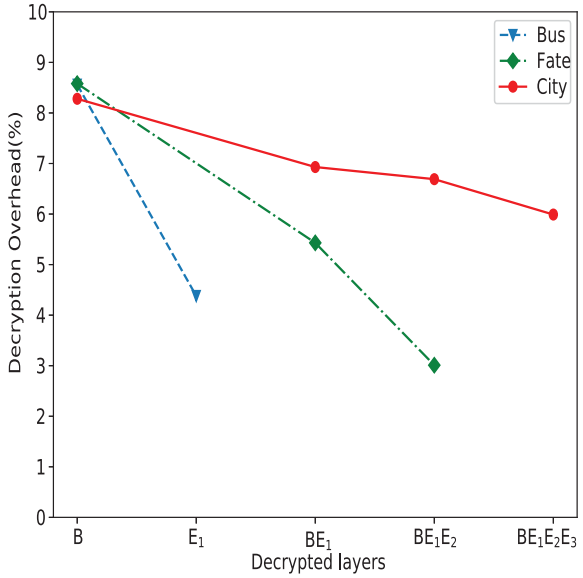


Fig. 12. Decryption overhead.

image after decrypting both the base layer and enhancement layer, and its resolution is CIF. Our scheme can guarantee the security of SVC bitstream without sacrificing the scalability.

B. Performance Analysis

1) Layered Encryption Effects: Taking the fate video sequence as a sample, we conducted a comparative experiment on the effect of layered encryption and decryption. As shown in Fig. 11, We statisticize the bitrate and peak signal-to-noise ratio (PSNR) of the video after decrypting the corresponding number of layers. The experimental results show that the bitrate and PSNR of the video increase with the number of decrypted layers, which means the quality of the video is improving. However, if only BL and EL2 are decrypted, the bitrate and PSNR are consistent with the BL layer, indicating that they have the same quality. That means the key of the Ith layer will not take effect until the $I-1$th is decrypted. So our proposed scheme can implement layered encryption and strict hierarchical control of SVC video.

2) Computational Cost: We use the ratio of decryption time to the sum of decoding time and decryption time to express as decryption overhead. Fig. 12 shows the decryption overhead for the test video sequences at different levels. It shows that as the number of decryption layers increases, the decryption load decreases. The enhancement layers are encrypted by stream cipher with low computational complexity. Fig. 12 indicates that the encryption algorithm of enhancement layers can accelerate the overall decryption efficiency.

As shown in Table IV, where ODS means original data size. ET is the encryption time and EET is the sum of the encoding time and the encryption time. Similarly, DT is the decryption time. DDT means the sum of the decoding time and the decryption time. EDS means encrypted data size and overhead is the ratio of the added data to the original data size.

TABLE IV
COMPUTATION COST AND FILE OVERHEAD

	ODS/byte	ET/EET	DT/DDT	EDS/byte	Overhead
Fate	162426	0.26%	3.01%	162450	0.014%
Bus	325143	2.57%	4.57%	325169	0.007%
City	3794921	0.59%	6.19%	3794951	0.0008%

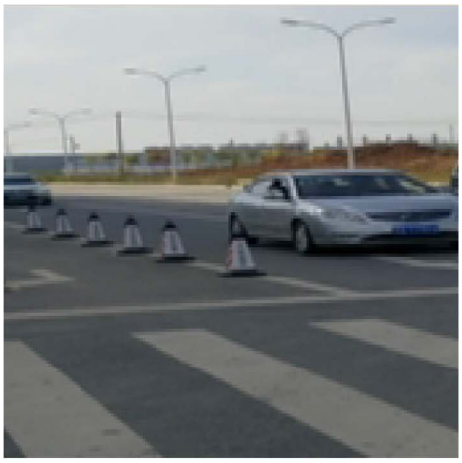
For all test video sequences, the computational cost of encryption is less than 2.57% and the computational cost of decryption is less than 6.19%, which are within an acceptable range relative to encoding time and decoding time. That means our scheme will not cause excessive time consumption when encoding and decoding. When the previous video segment is decoded and played, the next video segment has been decrypted and waited to be decoded and played. So there will be no delay in playback.

Comparing with the size of the video file before and after encryption, Table IV shows that the size of the video file has hardly changed after encryption. For all video sequences, the percentage of the increased data in original video data is less than 0.015%. All enhancement layers are encrypted with stream ciphers without changing the size of the file. While the size of the base layer is much smaller than the enhancement layer, and the size changed by AES encryption is almost negligible. Therefore, our scheme hardly needs to add other information to the code stream to generate additional bitstream data, which will not increase the requirement of the network bandwidth.

The experimental results show that the proposed scheme has acceptable computational complexity and less file storage loss. There is no additional burden on video transmission and storage. Therefore, our scheme will not have a negative impact on the real-time transmission of SVC videos.

C. Use Case

Our scheme can be applied to layered access control of IoT surveillance video. We can obtain a video stream from an edge



(a)

(b)

Fig. 13. Video sequence with different privileges. (a) Video sequence of low privilege. (b) Video sequence of high privilege.

camera for SVC coding and layered encryption. If the edge camera is powerful, the encoding and encryption process can be deployed on the edge device. Otherwise, deploy them on the server side. When users request videos, we continue to transmit SVC video segments to users for layered decryption and decoding. Depending on their keys, users can play videos of different quality or be denied access.

We code a section of traffic surveillance video in the SVC format for the experiment. As shown in Fig. 13(a), low-privilege users can only access the base layer. For example, ordinary drivers are restricted to view merely the outline of the surveillance video, who cannot see the license plate clearly while Fig. 13(b) shows that high-privilege users can access HD video to see the license plate, such as traffic management department staffs. For users, our scheme barely changed the size of surveillance videos which has no additional burden on transmission and decoding. For an operator, all surveillance videos need only to save one version for layered access control, which reduces storage overhead.

D. Comparison With Other SVC Encryption Schemes

Table V summarizes the performance comparison between our proposed scheme and the previous SVC encryption scheme, where complexity means computation complexity. Overhead is compression overhead. C&M are layered access control and key management.

TABLE V
COMPARISON BETWEEN OUR SCHEME AND OTHER SVC ENCRYPTION SCHEMES

Scheme	Security	Complexity	Overhead	C&M
[12],[13],[14],[15],[21]	Low	Low	No	No
[22]	Medium	Low	15%	No
[16],[17],[18],[19]	High	Medium/High	0.8%-2.8%	No
Our Scheme	Highest	High	0.015%	Yes

For security, [12]–[15] and [21] expose semantic content of video by setting all the signs be positive or negative [26], which means low security level while [22] is relatively higher than them. Li *et al.* [16], Hellwagner *et al.* [17], Stütz and Uhl [18], and Wei *et al.* [19] used an encryption algorithm technology for SVC bitstream, which has high security level. As far as our scheme, we encrypt the entire SVC bitstream, which uses the AES algorithm to encrypt the base layer and different encryption keys to encrypt each enhancement layer through stream cipher. Therefore, our scheme has higher security level.

For computation complexity, [12]–[15], [21], and [22] only need to perform XOR operations on the sign of motion vectors, intraframe modes, and nonzero coefficients with low computational complexity. For all NAL units of SVC bitstream, Li *et al.* [16] selected SPS, PPS, and IDR to encrypt with medium computational complexity. References [17]–[19] and our scheme require performing an XOR operation for every bit of SVC bitstream, thus that has high computational complexity. While the encryption load of our scheme is less than 2.57% and the decryption load is less than 6.19%, which are within an acceptable range.

For compression overhead, the file size before and after encryption is unchanged in [12]–[15], and [21] except of [22], which means no compression overhead. The compression overhead of [16]–[19] ranged from 0.8% to 2.8%. In our scheme, only the base layer would change its size after encryption, so the compression overhead is always less than 0.015%. Compared with other bitstream-oriented encryption schemes, the compression overhead is minimal and little additional bitstream data are added.

Compared with other SVC encryption schemes, our scheme implements the layered access control of SVC video and proposes a hierarchical key management scheme, which facilitates key management and user access control.

VI. CONCLUSION

In this article, we presented our novel hierarchical encryption and key management scheme designed to ensure the security of videos in H.264/SVC bitstream, as well as providing layered access control services. For the encryption, we used different cryptographic algorithms for the base layer and enhancement layers according to its significance; thus, achieving security and improved computational efficiency. Additionally, we proposed a hierarchical key management scheme according to the principle of the H-D wallet. The

key of the high-level user can derive all lower level keys for decryption but is not capable of obtaining higher level keys. In other words, users of any level only need to manage one key, and users without any key cannot access the video. The experimental results showed that the proposed scheme achieves a high security level, low compression overload, and acceptable computational complexity, in comparison to other SVC encryption schemes. In addition, our encryption and key management scheme provide the layered access control of SVC. The quality of the video content accessed by users depends on their authority, which can be applied to video surveillance systems of IoT.

REFERENCES

[1] C. V. N. Index, "Forecast and methodology, 2016–2021," White Paper, Cisco Public, San Jose, CA, USA, vol. 6, 2017.

[2] T. Ma, M. Ma, and F. Hu, "Scalable protection scheme for the H.264/SVC video streaming," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nanjing, China, 2017, pp. 1–6.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[4] M. Wien, H. Schwarz, and T. Oelbaum, "Performance analysis of SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1194–1203, Sep. 2007.

[5] J. Reichel, *Joint Scalable Video Model (JSVM) 2.0 Reference Encoding Algorithm Description*, ISO/IEC JTC1/SC29/WG11, N7084, Buzan, South Korea, 2005.

[6] A. I. Hentati, L. C. Fourati, and O. B. Rhaiem, "New hybrid rate adaptation algorithm (HR2A) for video streaming over WLAN," in *Proc. 6th Int. Conf. Commun. Netw. (ComNet)*, Hammamet, Tunisia, 2017, pp. 1–6.

[7] S.-W. Park, J. W. Han, and S.-U. Shin, "Secure service mechanism of video surveillance system based on H.264/SVC," in *Proc. 5th Int. Conf. Inf. Technol. Multimedia (ICIMU 2011)*, Kuala Lumpur, Malaysia, 2011, pp. 1–4.

[8] G. Bakar, R. A. Kirmizioglu, and A. M. Tekalp, "Motion-based rate adaptation in WebRTC videoconferencing using scalable video coding," *IEEE Trans. Multimedia*, vol. 21, no. 2, pp. 429–441, Feb. 2019.

[9] M. Vandana and H. Kallinatha, "Quality of service enhancement for multimedia applications using scalable video coding," in *Proc. 2nd Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Madurai, India, 2018, pp. 394–399.

[10] L. Hue, D. Duong, and X. HoangVan, "HEVC based distributed scalable video coding for surveillance visual system," in *Proc. 4th NAFOSTED Conf. Inf. Comput. Sci.*, 2017, pp. 314–318.

[11] R. H. Deng, X. Ding, Y. Wu, and Z. Wei, "Efficient block-based transparent encryption for H.264/SVC bitstreams," *Multimedia Syst.*, vol. 20, no. 2, pp. 165–178, 2014.

[12] Y. G. Won, T. M. Bae, and Y. M. Ro, "Scalable protection and access control in full scalable video coding," in *Proc. Int. Workshop Digit. Watermarking*, 2006, pp. 407–421.

[13] S.-W. Park and S.-U. Shin, "Combined scheme of encryption and watermarking in H.264/scalable video coding (SVC)," in *New Directions in Intelligent Interactive Multimedia*. Heidelberg, Germany: Springer, 2008, pp. 351–361.

[14] S.-W. Park and S.-U. Shin, "Efficient selective encryption scheme for the H.264/scalable video coding (SVC)," in *Proc. 4th Int. Conf. Netw. Comput. Adv. Inf. Manag.*, vol. 1. Gyeongju, South Korea, 2008, pp. 371–376.

[15] Y. Kim, S. H. Jin, T. M. Bae, and Y. M. Ro, "A selective video encryption for the region of interest in scalable video coding," in *Proc. TENCON IEEE Region 10 Conf.*, Taipei, Taiwan, 2007, pp. 1–4.

[16] C. Li, X. Zhou, and Y. Zhong, "NAL level encryption for scalable video coding," in *Proc. Pac. Rim Conf. Multimedia*, 2008, pp. 496–505.

[17] H. Hellwagner, R. Kuschnig, T. Stütz, and A. Uhl, "Efficient in-network adaptation of encrypted H.264/SVC content," *Signal Process. Image Commun.*, vol. 24, no. 9, pp. 740–758, 2009.

[18] T. Stütz and A. Uhl, "Format-compliant encryption of H.264/AVC and SVC," in *Proc. 10th IEEE Int. Symp. Multimedia*, Berkeley, CA, USA, 2008, pp. 446–451.

[19] Z. Wei, Y. Wu, X. Ding, and R. H. Deng, "A scalable and format-compliant encryption scheme for H.264/SVC bitstreams," *Signal Process. Image Commun.*, vol. 27, no. 9, pp. 1011–1024, 2012.

[20] M. N. Asghar, M. Ghanbari, M. Fleury, and M. J. Reed, "Sufficient encryption based on entropy coding syntax elements of H.264/SVC," *Multimedia Tools Appl.*, vol. 74, no. 23, pp. 10215–10241, 2015.

[21] S. Liu, S. Rho, W. Jifara, F. Jiang, and C. Liu, "A hybrid framework of data hiding and encryption in H.264/SVC," *Discrete Appl. Math.*, vol. 241, pp. 48–57, May 2018.

[22] G. B. Algin and E. T. Tunali, "Scalable video encryption of H.264 SVC codec," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 353–364, 2011.

[23] P. Wuille. (2012). *Bip32: Hierarchical Deterministic Wallets*. [Online]. Available: https://github. com/genjix/bips/blob/master/bip-0032. md

[24] C.-I. Fan, Y.-F. Tseng, H.-P. Su, R.-H. Hsu, and H. Kikuchi, "Secure hierarchical bitcoin wallet scheme against privilege escalation attacks," in *Proc. IEEE Conf. Depend. Secure Comput. (DSC)*, Kaohsiung, Taiwan, 2018, pp. 1–8.

[25] J. Reichel and H. Schwarz, "Wien M. joint scalable video model JSVM-19," Dept. ISO/IECJTC/SC29/WGll and ITU-T SGI6 Q.6, Joint Video Team, Geneva, Switzerland, Rep. JVT-V202, 2010.

[26] C.-P. Wu and C.-C. J. Kuo, "Fast encryption methods for audiovisual data confidentiality," in *Proc. Multimedia Syst. Appl. III*, vol. 4209, 2001, pp. 284–295.

Cheng Xu is currently pursuing the master's degree with the School of Computer Science, China University of Geosciences, Wuhan, China.

His research interests include streaming media transmission and scalable video coding.

Wei Ren (Member, IEEE) received the Ph.D. degree in computer science from Huazhong University of Science and Technology, Wuhan, China.

He is currently a Full Professor with the School of Computer Science, China University of Geosciences Wuhan, Wuhan. He was with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL, USA, from 2007 to 2008, the School of Computer Science, University of Nevada Las Vegas, Las Vegas, NV, USA, from 2006 to 2007, and the Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong, from 2004 to 2005. He has published more than 70 refereed papers, one monograph, and four textbooks. He has obtained ten patents and five innovation awards.

Prof. Ren is a Senior Member of the China Computer Federation.

Linchen Yu received the Ph.D. degree in computer science from Huazhong University of Science and Technology, Wuhan, China.

She is an Associate Professor with the School of Cyber Science and Engineering, Huazhong University of Science and Technology. Her research interests cover distributed computing and cloud computing.

Tianqing Zhu (Member, IEEE) received the B.Eng. and M.Eng. degrees from Wuhan University, Wuhan, China, in 2000 and 2004, respectively, and the Ph.D. degree in computer science from Deakin University, Geelong, VIC, Australia, in 2014.

She was a Lecturer with the School of Information Technology, Deakin University from 2014 to 2018. He is currently a Senior Lecturer with the School of Software, University of Technology Sydney, Sydney, NSW, Australia. Her research interests include privacy preservation, data mining, and network security.

Kim-Kwang Raymond Choo (Senior Member, IEEE) received the Ph.D. degree in information security from the Queensland University of Technology, Brisbane, QLD, Australia, in 2006.

He currently holds the Cloud Technology Endowed Professorship with the University of Texas at San Antonio, San Antonio, TX, USA.

Dr. Choo is a recipient of the UTSA College of Business Col. Jean Piccione and Lt. Col. Philip Piccione Endowed Research Award for Tenured Faculty in 2018, the IEEE TrustCom Best Paper Award in 2018, the ESORICS Best Research Paper Award in 2015, the Highly Commended Award by the Australia New Zealand Policing Advisory Agency in 2014, the Fulbright Scholarship in 2009, the Australia Day Achievement Medallion in 2008, and the British Computer Society's Wilkes Award in 2008. He was named the Cybersecurity Educator of the Year—APAC (Cybersecurity Excellence Awards are produced in cooperation with the Information Security Community on LinkedIn) in 2016, and he and his team won the Digital Forensics Research Challenge organized by Germany's University of Erlangen–Nuremberg in 2015. He is also a Fellow of the Australian Computer Society.