Article

Lightweight Cipher for H.264 Videos in the Internet of Multimedia Things with Encryption Space Ratio Diagnostics

Amna Shifa [1], Mamoona Naveed Asghar [1,2,*], Salma Noor [3], Neelam Gohar [3] and Martin Fleury [4]

[1] Department of Computer Science & IT, The Islamia University of Bahawalpur, 63100, Pakistan; amnashifa@yahoo.com
[2] Software Research Institute, Athlone Institute of Technology Ireland; masghar@ait.ie
[3] Department of Computer Science, Shaheed Benazir Bhutto Women University, Peshawar, Pakistan; dr.salmanoor@sbbwu.edu.pk, neelam.gohar@sbbwu.edu.pk
[4] School of CSEE, University of Essex, Colchester, Essex, UK; fleury.martin55@gmail.com
* Correspondence: masghar@ait.ie or mamona.asghar@iub.edu.pk or amnashifa@yahoo.com

Abstract: Within an Internet of Multimedia Things, the risk of disclosing streamed video content, such as arising from video surveillance, is of heightened concern. This leads to the encryption of that content. To reduce the overhead and lack of flexibility arising from full encryption of the content, a good number of selective-encryption algorithms have been proposed in the last decade. Some of them have limitations, in terms of significant delay due to computational cost, or excess memory utilization, or, despite being energy efficient, do not provide a satisfactory level of confidentiality, due to their simplicity. To address such issues, this paper presents a lightweight selective encryption scheme, in which encoder syntax elements are encrypted with the innovative EXPer (EXtended Permutation with exclusive OR). The selected syntax elements are taken from the final stage of video encoding that is during entropy coding. As a diagnostic tool, the Encryption Space Ratio measures encoding complexity of the video relative to the level of encryption so as to judge the success of the encryption process, according to entropy coder. A detailed comparative analysis of EXPer with state-of-the-art algorithms confirms that the EXPer provides significant confidentiality with a small computational cost and negligible encryption bitrate overhead. Thus, the results demonstrate that the proposed security scheme is a suitable choice for constrained devices in an Internet of Multimedia Things environment.

Keywords: Encryption Space Ratio; Entropy coding; H.264/AVC; Internet of Multimedia Things; lightweight cipher; selective encryption

1. Introduction

An Internet of Things (IoT) is a networked architecture [1], of which the Internet of Multimedia Things (IoMT) [2] is an emerging sub-set, integrating many devices and sensors at the Internet edge. In IoMT applications, video-surveillance devices might be deployed in various scenarios, such as within public transport management systems (managing buses, airplanes or road traffic), health management services (for patient or child monitoring), personal asset protection (within homes or construction sites) and many more [3]. The aim is to make these devices intelligent by allowing them to interact with each other, that is they become smart objects. Storage and later analysis of data [4] can be on remote cloud data centers. However, even more so than within the traditional Internet, the IoT architecture [5] has inherent security weaknesses of which this paper focuses on confidentiality. Furthermore, the adoption of multimedia rich content, within videos or images, has increased considerably in IoT environments, with the result that the Motion Pictures Experts Group (MPEG) is now standardizing audio-visual and other media data formats [6] as part of an IoMT. The devices in the IoMT, usually rely on Raspberry Pi and the CMOS platforms with limited computing and communication capabilities, hence, are not powerful enough for complex computations. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 1228; doi:10.3390/s19051228

adaptation of these limited resource devices in IoMT, particularly, in surveillance and monitoring systems are constantly increasing. Therefore, these require adequate security measures to keep the information secure. However, in an IoMT in which devices have limited resources in terms of processor power and memory and may well be battery powered, there is a requirement for simplified and computationally less complex ciphers. There is also a need to reduce the latency of communication by reducing the overhead arising from full encryption of all the contents. Because full encryption requires decryption at each intermediate point [8], for example if transcoding, splicing of content, adding logos or watermarks needs to take place, not only is there an added computational burden but there is a risk of key disclosure at unsupervised intermediate devices. Thus, full encryption is also an inflexible form of encryption. To cope with these challenges, the technology of lightweight cryptography has been utilized to provide efficient solution for securing information. Thus, this paper proposes a new lightweight stream cipher which is designed to be implemented easily on surveillance cameras operated with Raspberry Pi and the CMOS sensor platform.

This paper assumes that: confidentiality for video sensor networks [8] is integrated into the IoMT and that application-layer encryption is used. The alternative is to rely on any underlying end-to-end security protocol such as Transport Layer Security (TLS) (RFC 6176) with the overhead of full encryption and an associated Public Key Infrastructure (PKI) network of servers. Consequently, a lightweight encryption cipher, EXtended Permutation with exclusive OR (EXPer), is proposed, which combines statistically random output with speed of encryption, especially if only selected video syntax elements are encrypted. Those video syntax elements, encrypted as part of a Selective Encryption (SE) scheme, also preserve decoder format compliance, in the sense that the encrypted video can still be processed even if the selected syntax elements are not decrypted [9]. Notice also that if the authors' contribution in [10] is applied, keys may alternatively be embedded within the encrypted video itself through the joint application of steganography and cryptography, though that aspect is outside the scope of the current paper.

To judge the protection afforded by EXPer, the Encryption Space Ratio (ESR) is calculated as the ratio of the encrypted bits to all the bits of a compressed video bitstream. The ESR has previously been used as one way to judge the effectiveness of SE. In a Variable Bit-Rate (VBR) stream, the bitrate is dependent on the various configuration factors but if these are held constant across a number of test videos then both the spatial and temporal coding complexity change according to the content. For example, the presence of high spatial frequencies or textures within objects increases the coding complexity and likewise rapidly moving objects across video frames increases the temporal complexity. The Quantization Parameter (QP) also affects the bits allocated, with a low QP implying a higher compressed bitrate. The SE bitrate depends on the syntax elements that are selected for encryption. These syntax element bits may be fully encrypted, typically by a block cipher such as the Advanced Encryption Standard (AES) [11] operating in a streaming mode such as Cipher Feedback (CFB), but herein by EXPer for reduced latency. As an example of using the ESR to judge the effectiveness of SE, in [12], nine test videos each encoded with QPs at 16, 24, and 28 (High Efficiency Video Coding (HEVC)) with an ESR ranging from 12.42% to 20.11% and an average of 16.54%, which was judged to be sufficient. In [13] also, the ESR was calculated for a QP of 18 and found to vary from 16.96% to 20.08% depending on the content of six benchmark videos, when using a SE scheme for HEVC. In [13], which provides an analysis of different influences on the ESR such as the impact of background or HEVC codec profile, the ESR was examined for its diagnostic value. However, the following research questions arise from previous SE-based studies:

Q1. Are those encryption schemes computationally efficient (in terms of execution/encoding time) enough to employ in an IoMT communication environment?

Q2. Is the analyzed ESR is effective enough for applying SE to visually secure the videos encoded with one or other of the two common entropy coders in common codec use, i.e. Context Adaptive Variable Length Coding (CAVLC) and Context Adaptive Binary Arithmetic Coding (CABAC) (see Section 3.2)?

The focus of this paper is to answer these questions by experimentation. In the case of full (sometimes known-as) naïve encryption, the complete video is encrypted. Therefore, the encryption

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at Sensors 2019, 19, 1228; doi:10.3390/s19051228

overhead and the space ratio is at a maximum, which causes a bit-rate overhead too. Herein, such weaknesses are addressed by proposing a complete lightweight security scheme for IoMT applications on a standardized H.264/AVC encoder for constrained surveillance devices. Notice that HEVC is currently too resource intensive to be used in an IoMT environment, see [14]. In the proposed scheme, SE is applied through proposed encryption algorithm EXPer, over identified ESR for effective visual protection. The point wise contribution of the paper is given in the succeeding sub-section 1.2.

The remainder of this paper is organized as follows. In Section 2, prior efficient lightweight schemes proposed by researchers of IoT are discussed. Section 3 presents the proposed lightweight cipher scheme EXPer, the adopted SE methodology, and diagnostics through ESR. Section 4 describes the promising results over tested videos. Section 5 is a comparative analysis of EXPer with state-of-art ciphers, including Advanced Encryption Standard (AES) [11]. Finally, Section 6 rounds off by considering the implications for those planning IoMT video applications with a concern for confidential video content.

1.2. Context

SE already has a potential role in consumer electronics applications [15] and also can support interoperability [16], when multiple encryptions of the same video stream are transported. Alternatively, Region-of-Interest (ROI) encryption of some parts of a video frame such as the face or people within a frame [17][18] may reduce the encryption overhead. However, ROI encryption is application specific, while SE potentially offers a more general solution. Compared to full (or naïve) encryption, both SE and ROI encryption can reduce computational and bitrate overhead [19]. SE may be carried out on the most significant information (as regards distortion) at a choice of different stages of the codec, such as on the original pixels, the transform coefficients, the quantization indexes, the bit-planes, the entropy coder, or the final output bitstream [20]. However, some forms of encryption alter the video statistics, resulting in the issues of encryption bitrate overhead and format compliance at the decoder. Applying encryption at the entropy coding stage minimizes these problems [17] [18], which is why this form of encryption is chosen for this paper.

Entropy coders are a feature of standardized hybrid video encoders [21]. H.264/Advanced Video Coding (AVC) [22] and its Scalable Video Coding (SVC) extension [23-25] employ the same entropy coding modes: variable length coding (VLC) or binary arithmetic coding (BAC). Both of these modes operate in a context-adaptive (CA) manner, leading to the names CAVLC [26] and CABAC [27] entropy coders. Within H.264/AVC either CABAC or CAVLC entropy coders can be selected, as the two coders trade-off computational complexity against compression efficiency. The High Efficiency Video Coding (HEVC) [28] CABAC encoder is a slightly modified version of the H.264/AVC CABAC encoder and, thus, entropy-integrated SE can be configured [29] to work with either codec. However, using HEVC for an IoT is questionable owing to its high computational complexity, except possibly when coding takes place at a cluster head with maximum energy [30].

The H.264/AVC codec is selected for implementation of lightweight encryption because, in both the CCTV industry and for smart monitoring in IoT, surveillance devices (cameras) mostly operate on microprocessors, especially the Raspberry Pi (RasPi), which only supports video compression in the H.264/Advanced Video Coding (AVC) format [31]. In an IoT, the RasPi is an economical and privileged platform because it offers a complete Linux server on a tiny platform. To the best of the authors' knowledge, the SE utilized with ESR diagnostics is a novel contribution in securing IoMT communication. The contributions of the paper are:

1. A joint crypto-compression scheme is implemented on the selected video syntax elements in the entropy engine of H.264 encoder. The selective encryption is applied by keeping in mind the IoT constrained devices operated with Raspberry pi and CMOS camera sensors. The compatibility of the proposed cipher was prior tested on the Raspberry Pi cameras.
2. Selection of careful video syntax elements for applying selective encryption, which do not crash the video decoders.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at Sensors 2019, 19, 1228; doi:10.3390/s19051228

3. The Encryption Space Ratio (ESR) for two entropy coders, CAVLC and CABAC, is evaluated over 10 tested videos. ESR is used as a tool for applying efficient SE to visually secure the videos, directly recorded through IoMT constrained RasPi cameras.
4. For effective visual degradation of videos, it is found that the ESR estimated for CABAC is less than CAVLC, so CABAC entropy coder is suitable for IoT based camera devices.
5. A single round lightweight Cipher "EXPer" (with five sequential steps of bit level permutation and XOR) is proposed for the H.264/AVC videos in IoMT.
6. Series of experiments are done over H.264 encoder (JSVM reference software) (Section 4 and 5) with the proposed lightweight cipher on estimated ESR, and in comparison with two implemented state-of-the-art algorithms (XOR, AES). Therefore, three different ciphers are extensively tested (from different angles) on multiple videos with varying color and motion characteristics. The perceptual strength of implemented ciphers is compared through different quality metrics and their computational efficiency is evaluated in terms of execution time on tested videos.
7. Adopted methodology with experiments proved that the proposed lightweight security scheme (EXPer applied on calculated ESR for selective encryption) is secure against key guessing, perceptual, statistical and inference attacks in IoMT environment.

2. Related Studies

The heterogeneity of the communication technologies across an IoMT, deriving from the assortment of devices, sensors, and protocols, is a cause of security concern. Messages that are transmitted from smart objects will usually be stored and forwarded through several nodes (devices, video sensors, video relay or intermediate devices …) to reach their endpoint [7]. An endpoint might be a message sink or base station (BS), which passes data to the next layer of the architecture, as occurs in traditional sensor networks when a sink communicates over a satellite link. Figure 1 shows end-to-end communication of multimedia sensor networks within an IoT. Again, there is a need to preserve confidentiality over that link but there is also a need to optimize energy consumption and reduce latency.

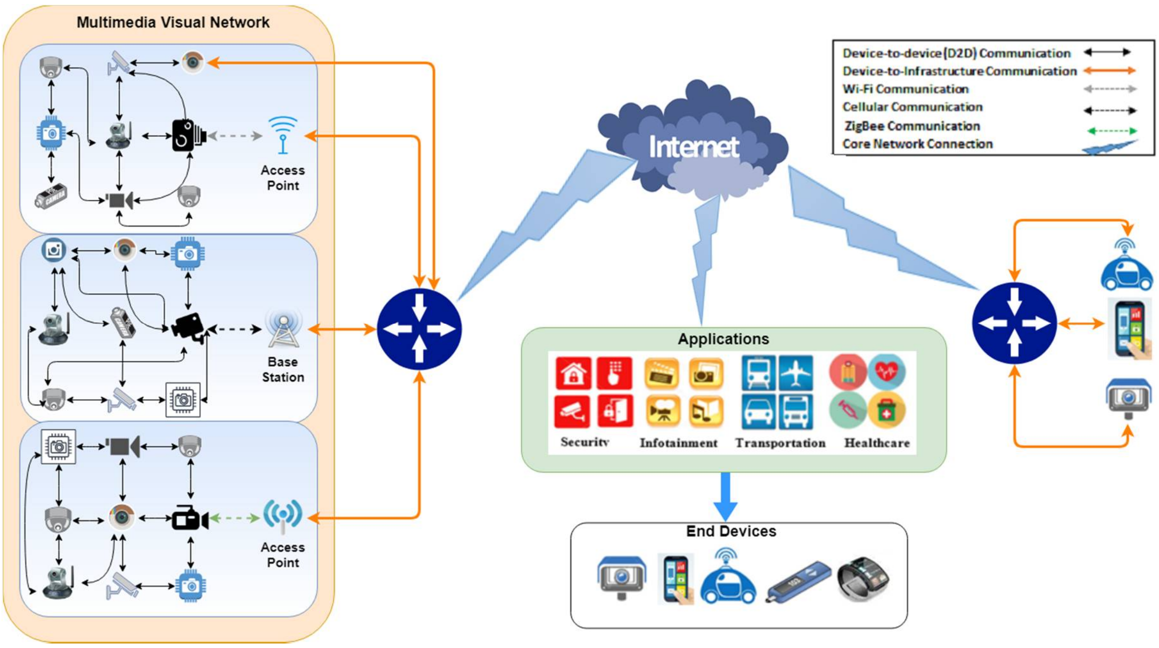


Figure 1. End-to-end communication over multimedia sensor networks in an IoT

To ensure confidential communication of information in an IoT, many schemes have been proposed, schemes which employ existing state-of-art encryption algorithms [32-38]. A summary of the proposed encryption schemes for IoT with standard cipher algorithms is presented in Table 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at Sensors 2019, 19, 1228; doi:10.3390/s19051228

Besides encryption, researchers have also proposed authentication schemes for IoT with existing standardized algorithms. Lee et al. [39] proposed a lightweight authentication protocol for RFID systems. In the proposed protocol, privacy protection and anti-counterfeiting is achieved by an encryption algorithm based on XOR manipulation. Mahalle, et al. [40] proposed Identity Authentication and Capability-based Access Control (IACAC). The proposed scheme provided both authentication and access control for an IoT. However, the proposed scheme presents extra overhead, due to its key-management procedures. The authors [41] employed encryption and hash algorithms in their proposed solution to achieve confidentiality and message integrity in an IoT. However, their solution fails to deal with large amounts of multimedia data because of the proposed encryption algorithm's ability to encrypt only 64 bits per block; hence, it suffers from slow operation.

Table 1. Summary of recently proposed security schemes for IoT by using existing standardized encryption algorithms

Proposed Scheme	Year	Paltform	Algorithm Used	Strengths	Limitations
El Assad S, Farajallah M [34]	2016	FPGA card or an ASIC/	Diffusion and Chaotic Map (2D cat map)	Efficient and provides a high level of security, resistance to known plaintext and chosen plaintext attacks	Requires huge memory capacity
Al-Salami et al. [35]	2016	Smart Home	Identity-based encryption (IBE), StatefulDiffie-Hellman (DH) Encryption, Private Key Generator (PKI)	Provides favorable computational and communication efficiency	Overhead of handling private key generator (PKG)
Yao et al. [36]	2015	IoT	Attribute-based encryption (ABE), Elliptic Curve Cryptography(ECC), Diffie–Hellman (DH), Elliptic Curve Decisional Diffie–Hellman (ECDDHP)	Improved execution efficiency and reduced communication costs	Poor flexibility in revoking attributes and weak scalability
Xin M [37]	2015	IoT	MD5, Elliptic Curve Cryptography (ECC) and AES algorithm,	Improved security and performance	Increase the complexity and reduce execution speed
Prasetyo et al. [38]	2014	FPGA module/ Constraints devices	Blowfish algorithm	Better security and reduced total encryption time	Larger key length requires more resources and suffers from error propagation

Although, the encryption schemes given in Table 1 provide higher security, the authors did not consider multimedia content when evaluating the performance of their proposed schemes. Moreover, traditional encryption algorithms, such as AES and Triple Data Encryption Standard (DES) encryption, as used in the proposed schemes are inefficient for an IoMT because of their computationally intensive nature. Hence, those schemes appear to be unsuited to the requirements of real-time IoMT applications, due to their relatively high bitrate overhead, computational overhead, and bandwidth utilization. Consequently, lightweight encryption algorithms are required to alleviate these overheads for low-cost, low-power devices. Recently, there has been much interest shown by researchers and standardization bodies in designing lightweight algorithms for secure end-to-end communication in an IoT. All cryptographic algorithms are based on three principles 1) Substitution 2) XOR and 3) Permutation. Thus, the newly proposed algorithms by other researchers are also based on these principles. Recently, the authors of [47] proposed one round cipher (implemented on static images) for IoMT in which substitution and permutation principle have been selected for the encryption. However, the substitution is considered resource expensive and should be used with caution over videos, especially for the resource-limited devices in the IoMT. Thus, in this study, the lightweight cipher is employed with the single round of five (3 XOR and 2 Permutation) sequential steps over videos. It has been tested that the substitution rounds cannot be efficiently computed over videos in the resource-limited device used in IoMT. An overview of some recently proposed lightweight encryption algorithms in comparison with our proposed encryption algorithm for IoMT communication is given in Table 2.

Table 2. Overview of recently proposed lightweight ciphers for constrained devices

Algorithm	Target Multimedia	Structure	Target Devices/ Platform	Cipher Type	Key Size (Bits)	No.of Rounds	Secure against Attacks
PRESENT (2007) [42]	General purpose	Substitution-Permutation Network (SPN)	RFID tags and sensor	Block	80,128	31	1-Side-channel attacks, 2-Differential attacks
Hummingbird2 (2012) [43]	General purpose	Hybrid	Low-end controllers, RFID tags, wireless sensors, smart meters	Hybrid	256	4	1-Related-key attack, 2-Side-channel attacks
TWINE (2013) [44]	General purpose	Type-2 generalized Feistel network (GFN-2)	Micro-controller and high-end CPU.	Block	80/128	32	Meet-in-the-middle attacks
PRIDE (2014) [45]	General purpose	Substitution permutation network	8-bit micro-controller	Block	128	20	1-Meet-in-the-Middle attacks, 2-Differential attacks
Lightweight chaotic image encryption algorithm (2018) [46]	Image	Chaotic map	32-bit microcontroller and real-time embedded applications	Block	128	Not specified	Not specified
One round encryption algorithm (2018) [47]	Image	Substitution permutation network	Multimedia IoT	Stream	512	1	Not specified
Proposed EXPer (2019)	**H.264 Videos**	**bit-wise Permutation and XOR over CABAC encoded Syntax elements (Bins)**	**Multimedia IoT**	**Stream**	**128**	**1**	**1-Perceptual Attacks, 2-key guessing Attacks, 3-Statistical Attacks, 4-Inference Attacks**

Likewise, to avoid extra computational overhead and bitrate control, chaos theory is also utilized to implement the encryption process for IoMT systems [48][49]. Chaos theory has proved attractive because of its simplicity and statistical qualities leading to randomized output. Generally, chaotic algorithms are based on a chaotic map and s-box substitution, with multiple rounds to create randomization. However, in the substitution process, multiple rounds to create the desired random output increase the execution time. In fact, in [50] doubt has been cast upon the computational gain from employing chaotic encryption, compared to traditional block-based encryption such as through AES. Indeed, statistical tests often used to verify the confidentiality of chaotic encryption fail to highlight known insecure encryption algorithms, casting doubt on the claimed security properties.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 1228; doi:10.3390/s19051228

In general, researchers have proposed ciphers for general-purpose applications and do not consider the specific ESR required to provide effective visual protection especially in images and videos.

3. Materials and Methods

This section explains the approach of the proposed security scheme for devices in IoMT environment. The security scheme is comprised of four components:

1. Selection of Syntax Elements for two Entropy Engines i.e. CAVLC and CABAC.
2. EXPer, an innovative lightweight cipher based on a combination of eXclusive OR (XOR) and bit-level permutation rounds with three different 128-bit keys.
3. ESR is calculated according to the selected syntax elements of CAVLC and CABAC, as a way of diagnosing the effective visual protection. (Section 4)
4. SE is applied by utilizing EXPer, according to the guidance given by ESR. (Section 4)

3.1. Syntax Elements Selection of Entropy coders

There are two forms of entropy engines for efficient compression in H.264/AVC video encoder, CAVLC [21] and CABAC [22]. Both CAVLC and CABAC are a lossless form of coding (after earlier lossy encoding stages of the hybrid codec) in which there is a tight data dependency between elements in the output bitstream. CAVLC employs the concatenation of Unary and Exp-Golomb coding for a number of parameters, such as macroblock (MB) type (i.e. the prediction method — inter or intra), coded block pattern (CBP) (which records which blocks within an MB contain Non-Zero (NZ) Transform Coefficient (TC) residuals), delta QP, reference frame index, and Motion Vector Differences (MVDs). Quantized transform block coefficients (residuals) are VLC coded after extraction (normally through zig-zag scanning) from a block. CABAC gains in efficiency over CAVLC as syntax elements are first converted to a binary format. This allows binary arithmetic coding to be utilized. Arithmetic coding leads to sub-integer probability estimation (unlike CAVLC) but is computationally expensive.

In this paper, for the convenience of implementation the focus is not on a newly proposed SE encoder technique. Therefore, we have employed our previous SE schemes, e.g. as reported in [51], with one newly identified parameter for enhanced visual protection of videos. For the convenience of new researchers, we give simple names to these selected parameters, i.e. (1) motion, dealing with the movement of objects in videos including camera zooming etc. (2) texture data for pixels information, and a new parameter (3) difference of Quantization Parameters (deltaQP). Their details can be found in [26] and [27].

It is also worth mentioning here that these mentioned three types of parameters are produced only from residual information, as presented to the final entropy coding stage of a hybrid encoder and are proven to be format compliant in experiments. Textural residuals are taken from both homogeneous and heterogeneous areas within a video. For motion encryption, the arithmetic signs of motion vector difference (MVD) are encrypted while textural syntax elements are different for both CAVLC and CABAC (given in the top grey box of Figure. 2), and the absolute values of dQP are selected for encryption. The selected syntax elements of CABAC will be referred as *Bins* in the subsequent sections of this paper. A block diagram of the proposed security scheme is given in Figure 2.

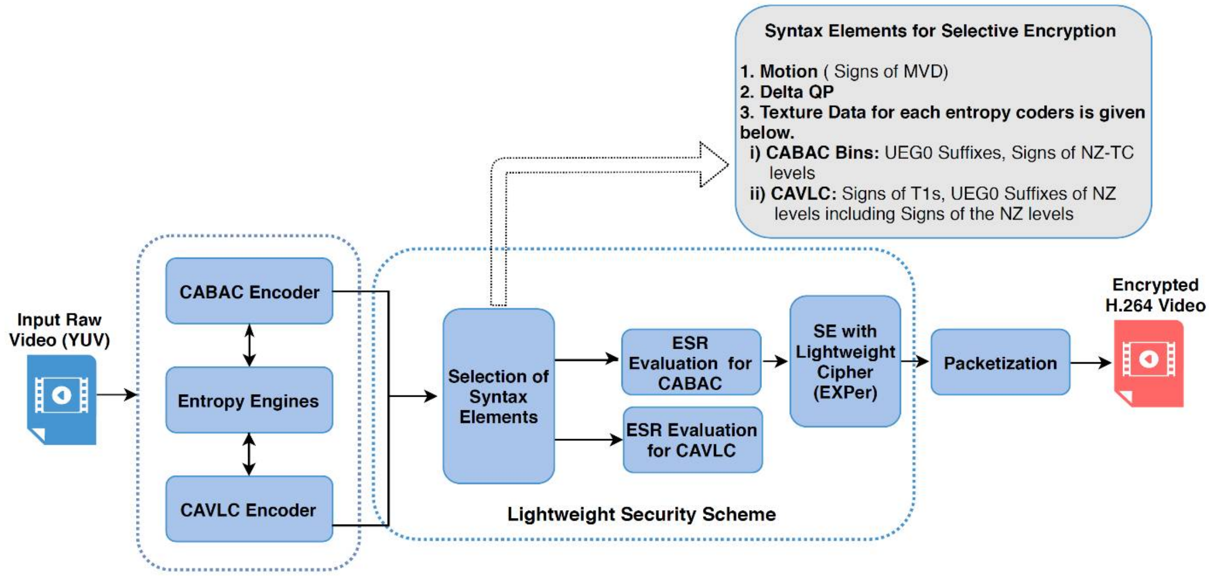


Figure 2. Block diagram of the Proposed lightweight security scheme

3.2. Lightweight Cipher

The proposed lightweight cipher, EXPer, provides both diffusion and confusion primitives, through XORing and permutation. Because the normal substitution process (in mainstream ciphers such as AES) is computationally intensive, so it is not included in the proposed cipher. Moreover, the different forms of XOR are also not computed over videos, because their adoption can be effective for a computation over single image rather than a whole video with large quantity of surveillance frames. Although permutation is applied on bit-level within selected byte to effectively secure the camera captured videos.

EXPer encryption consists of five steps/stages with a single iteration over those steps. In each step, XOR is performed using a secret key (k_s), and the bit-wise permutation by the shift operation. Permutation is performed with a randomly-selected offset value (V) ranging from 1 to 8. Additionally, permutation is performed on the output from the previous stage to approach statistical randomness with a reduced computational complexity. The permutation is applied on the bit-level, re-orders the bits within byte which is not easy to cryptanalyze for large volume of video data.

The symmetric secret keys: secret_key (k_s), sub_key1 (k_1), sub_key2 (k_2) are dynamically generated at run-time for the input bitstream, by using a Pseudo-Random Function (PRF). To keep the procedure simple, three dynamic keys are generated per video sequence and stored in a registry. Key security can be enhanced by using any standardized key management scheme [52]. Additionally, three 128-bit secret keys have been utilized with XOR operation. The key space greater than 2^{100} is considered resilient to key guessing or brute force attacks over keys [53]. Furthermore, the selective encryption on the selected syntax elements within large volume of videos data has proved to be strong, hence cannot be cryptanalyzed [54]. Moreover, the selected offset value will permute the bits within the byte only. We consider the proposed algorithm to be a stream cipher because of byte level encryption. In fact, the encrypted bitstream is obtained by combining the elements of the plaintext bitstream bit-by-bit with secret keys.

Working of EXPer:

The five steps of the proposed algorithm are discussed in more detail below:

Step 1: In the first step, the input bitstream is encrypted by performing an XOR operation with the 128-bit secret key (secret_key). Let X be the chosen *"Bins"*, which are the syntax elements of the CABAC entropy coder. Then, the secret key, k_s, is XORed with *X*. Mathematically:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1
Peer-reviewed version available at Sensors 2019, 19, 1228; doi:10.3390/s19051228

$$f(k_s)=X \oplus k_s \quad (1)$$

$$X \oplus k_s = X' \quad (2)$$

where $\oplus$ is the XOR operator, X is any input bitstream, and X' is the resulting output bitstream.

Step 2: In the second step, re-ordering of bits within selected bytes of syntax elements is performed by applying a permutation to previously encrypted output X' derived from f(ks). The elements of X" are cyclically shifted initially by offset value v_i using a right-shift operator. The right-shift operator shifts a bit pattern to the right by an offset value:

$$f(k_s) \mapsto f(k_s+v_i) = (x1, x2, x3\ldots) \mapsto (x2, x3, x4\ldots) \quad (3)$$

where "↦" is a transformation symbol. It this case it signifies transforming X' into the X'' bitstream through a right-shift of the bits of X'=(x1, x2, x3…xn), to become X''= (x2, x3, x4…..xn). The vi denotes an offset value and V represents a finite set of possible offset values that can be expressed as V= {1, 2, 3,...,m}, where n is the number of possible offsets. The initial value of offset vi can be selected dynamically to attain greater security. Mathematically,

$$X' >> v_i = X'' \quad (4)$$

where >> is a shift operator.

Step 3: In the next step, the resulting output of the previous step, which is X'', is again transformed by an XOR operation with the 128-bit sub-key1(k_1), as:

$$X'' \oplus k_1 = X''' \quad (5)$$

As already mentioned, $k_1$is derived from k_s.

Step 4: In the fourth step, the previously encrypted output X''' is permutated again with offset value v_j, again by a right-shift operator:

$$X''' >> v_j = X'''' \quad (6)$$

while $v_j = (v_i+2)$ (7)

Step 5: In the final step, the resulting bitstream, X'''', is XORed with the 128-bit sub_key2 (k_2),to produce encrypted bitstream E_{output}:

$$X'''' \oplus k_2 = E_{output} \quad (8)$$

The proposed algorithm is simple and, thus convenient to implement even on videos directly taken from RaspPi Cameras. The pseudocode and flowchart in Figure 3 demonstrates the simplicity of a software implementation with encryption and decryption rounds.

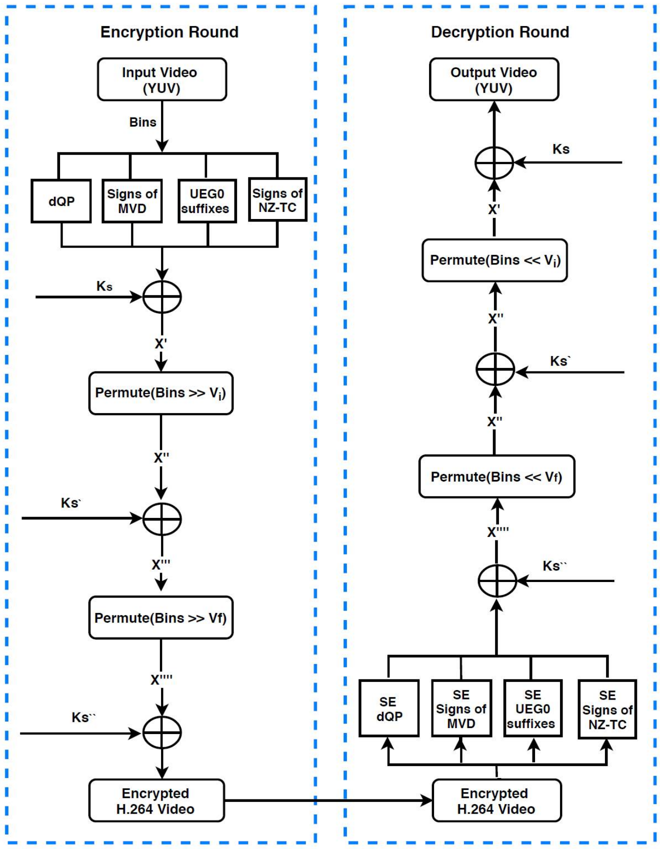
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 1228; doi:10.3390/s19051228

Pseudocode of EXPer

1. Input: Plaintext video (referred as PV).
2. Bins: uiSigns (MVD & NZ-TC), UEG0 Suffixes and dQP
2. Generate: Secret key (ks), sub_key1 (ks\`), sub_key2 (ks\`\`));
Keys are generated with a PRF and stored.
3.EXPerEnc for each selected Syntax Element:
{
4. Bins = Bins ^ ks;
5. Bins = Bins >> vi; suppose vi=3
6. Bins = Bins ⊕ (ks\`);
7. Bins = Bins >> vj; where vj=(vi+2)
8. Bins = Bins ⊕ ks\`\`;
}
9. Output: Encrypted video bitstream (referred to as EV)

(a)



(b)

Figure 3. (a) Pseudocode and (b) the graphical flowchart of proposed EXPer Cipher

4. Experimental Results and Discussion

In order to evaluate the performance, experiments were performed on ten well-known [55] test videos with varying characteristics, such as slow/fast motion and light/dense colours. The tested videos configurations were based upon Common Intermediate Format (CIF) (352 × 288 pixels/frame) at 30 fps, 4:2:0 chroma sampling, IBBP… Group of Pictures (GoP) frame structure and an intra-refresh period of length 16, with H.264/AVC. The videos are evaluated on different QP values. (The H.264/AVC the range of QPs is from 0 to 51, corresponding to higher compression with lower QPs). All experiments were performed on a 64-bit operating system with 2.30 GHz Core i5-6200U processor and 8 GB RAM. The algorithm was developed using the C/C++ programming language by modifying the JSVM reference software with a single layer [56].

4.1. Calculation of ESR for Entropy coders

Before applying SE with EXPer on test videos, the focus of this paper is to analyze the ESR of two entropy coders over which the SE is applied. ESR is basically the amount of data within each video (calculated in terms of percentages) over which the SE produces acceptable visual protection results (see also Section 1's introduction to ESR). The ESR percentage is directly proportional to the computational cost of applying SE over videos i.e. more ESR, more computational cost for SE and vice versa. The ESR for videos calculated on the bases of selected syntax elements (as specified in Section 3.2). The tested videos are listed in Table 3 and configured as described at the beginning of this Section.

Taking the ESR percentages for CAVLC first in Table 3, it is apparent that there is a considerable content dependency, probably linked to the spatial complexity of individual video frames and the temporal complexity, due to the level of motion activity within sequences. The ESR percentage of motion elements is much below that of the texture ESR, i.e. that arising from spatial complexity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 1228; doi:10.3390/s19051228

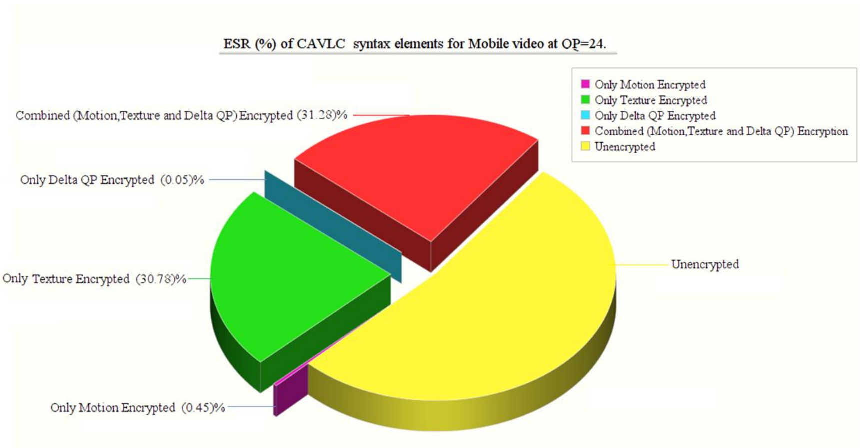
However, from the observations of Section 6.1, the ESR value calculated for motion parameter SE elements alone is insufficient in itself to guarantee encryption confidentiality. However, the ESR value for motion and texture SE elements, when those syntax elements are derived from CAVLC can be considerable with a maximum ESR of 29.27% for the Flower test video. It is also the case, that despite the view that SE syntax elements/parameters can be chosen so that in a statistical sense there is little impact on the bitrate overhead, in fact, from experimental evidence, the encryption ratio appears to be considerable.

Compared with SE of CABAC syntax elements in Table 3, the ESR value is considerably lower for CABAC, with an average (arithmetic mean) of 12% for CAVLC and 7.5% for the CABAC encoded test videos. Additionally, the maximum encryption ratio drops to 14.61% for CABAC and for Mobile rather than the Flower video. Given that CABAC already has an advantage in terms of compression efficiency (refer back to Section 3), so CAVLC must be adopted with caution for IoT applications. This is unfortunate given the reduced computational overhead arising from CAVLC and its inclusion in H.264/AVC's Baseline-type profiles. The pictorial comparison of ESR calculated for CAVLC and CABAC is illustrated in Figure. 4.

It is also worth noticing in Table 3 and Figure. 4 that for all tested samples, CAVLC produces more texture information than CABAC. This property consequently provides more texture ESR for encryption and, hence, produces more computational overhead than CABAC. For this reason, the EXPer experiments were performed with ESR on CABAC in the next section.

Table 3 Comparative ESR (%) of CAVLC and CABAC syntax elements for sample videos at QP=24.

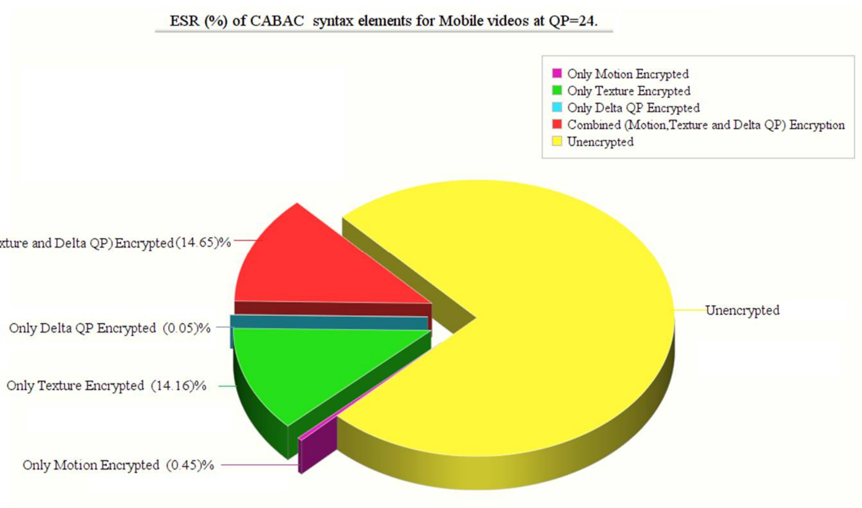
Sr. #	Videos (CIF)	File size(MB)	Encryption Ratio (%)					
			Only Motion Encrypted		Only Texture Encrypted		Combined (Motion and Texture) Encryption	
			CAVLC	CABAC	CAVLC	CABAC	CAVLC	CABAC
1.	Bus	21.7	0.55%	0.56%	23.42%	10.72%	23.97%	11.27%
2.	Coastguard	43.5	0.32%	0.32%	21.52%	10.10%	21.84%	10.41%
3.	Container	43.5	0.07%	0.07%	6.94%	3.17%	7.01%	3.24%
4.	Crew	43.5	0.45%	0.45%	12.57%	7.57%	13.01%	8.02%
5.	Flower	36.3	0.56%	0.57%	28.71%	13.57%	29.27%	14.14%
6.	Football	37.7	0.52%	0.52%	25.79%	11.30%	26.30%	11.81%
7.	Foreman	43.5	0.30%	0.30%	11.61%	5.05%	11.92%	5.35%
8.	Hall	36.3	0.10%	0.10%	11.11%	4.11%	11.21%	4.22%
9.	ICE	34.8	0.34%	0.34%	6.92%	2.87%	7.26%	3.22%
10.	Mobile	43.5	0.45%	0.45%	30.78%	14.16%	31.23%	14.61%



(a) ESR calculated for CAVLC syntax elements

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 1228; doi:10.3390/s19051228



(b) ESR calculated for CABAC syntax elements

Figure 4. Comparison of calculated ESR over (**a**) CAVLC and (**b**) CABAC Syntax Elements of Mobile video

4.2. Performance of EXPer

For the evaluation of results with EXPer, the experiments were performed on several test video sequences with the selected parameters such as those based on motion, texture, delta QP, and together with their combinations. The SE is applied with CABAC on all tested videos because CABAC is more compression efficient (refer back to Section 3) and produces less ESR as compared to CAVLC. Another reason for choosing CABAC is that the encryption of the by-pass syntax elements in CABAC does not affect the context models and encrypting at the entropy coder stage does not affect bitstream compliance at the decoder, which is why by-pass CABAC syntax elements are more appropriate for SE in IoMT. Table 4 shows the calculated ESR (ratio) for CABAC syntax elements.

The visual results with the proposed encryption algorithm EXPer on CIF video sequences Mobile, ICE and Stefan are presented in Figure. 5, which imply that sufficient confidentiality is achieved without generating encryption overhead. The ESR with CABAC for ten videos is depicted in Table 4. The ESR of delta QP is only 0.04% while 0.34% and 6.92% with only motion and texture respectively for the ICE video. The ESR for all parameters for ICE video is 4.25% with the CABAC coder, which is 95.75% less than the data for naïve encryption. The average ESR for test videos with all parameters combined (i.e. motion, texture and delta QP) is 8.69% which is minimal and can be adopted by IoT devices.

Table 4. ESR in terms of percentage over which SE applied on CABAC bin-strings

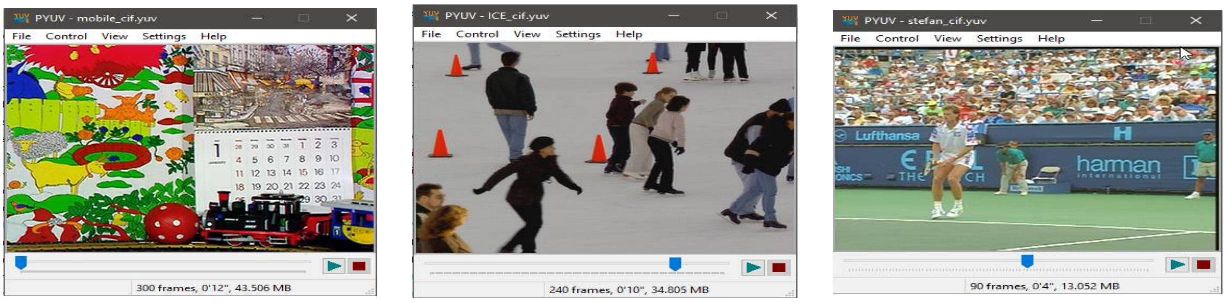
Sr.#	Videos (CIF)	File size (MB)	ESR (%)						
			Only Motion	Only Texture	Only Delta QP	Both Motion and Delta QP	Both Texture and Delta QP	Both Motion and Texture	Motion, Texture and Delta QP
1.	Bus	21.7	0.56%	10.72%	0.06%	0.62%	10.78%	11.28%	11.34%
2.	Coastguard	43.5	0.25%	4.46%	0.04%	0.29%	4.50%	4.71%	4.75%
3.	Container	43.5	0.32%	10.10%	0.02%	0.34%	10.12%	10.42%	10.44%
4.	Crew	43.5	0.07%	3.17%	0.08%	0.15%	3.25%	3.24%	3.32%
5.	Flower	36.3	0.45%	7.57%	0.05%	0.50%	7.62%	8.02%	8.07%
6.	Football	37.7	0.57%	13.57%	0.10%	0.67%	13.67%	14.14%	14.24%
7.	Foreman	43.5	0.52%	11.30%	0.03%	0.55%	11.33%	11.82%	11.85%
8.	Hall	36.3	0.30%	5.05%	0.03%	0.33%	5.08%	5.35%	5.38%
9.	ICE	34.8	0.10%	4.11%	0.04%	0.14%	4.15%	4.21%	4.25%
10	Mobile	43.5	0.46%	12.76%	0.04%	0.50%	12.80%	13.22%	13.26%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 1228; doi:10.3390/s19051228

Furthermore, as previously mentioned it is also worth noticing that the ESR with delta QP is comparatively lower than the ESR of only motion or only texture parameters for all tested videos. Thus, the ESR of delta QP combined with motion (0.50% for Mobile video and 0.14% for ICE video) or delta QP combined with texture (12.80% for Mobile video and 4.15% for ICE video) is also less as compared to the encryption ratio with combined motion and texture (13.22% for Mobile video and 4.21% for ICE video). The important point to note here is that the SE on the absolute values of dQP is not possible with complex cipher algorithms, because the number of rounds in complex ciphers makes these values out of range which destroys the format compliance and compression efficiency of the bitstream, and consequently crashes the decoder [48].

However, in this paper, visual results in Figure 5 (a4, b4, c4) show the effectiveness of EXPer algorithm, as the SE on dQP syntax elements is implemented in a way that their absolute values do not go out of range and, as a result, the compression efficiency and format compliance of videos are both maintained. This format compliance cannot be achieved through the AES algorithm. Overall, the results in Figure 5 indicate that EXPer encryption provides sufficient visual protection with a lower computational and encryption cost than conventional AES encryption.



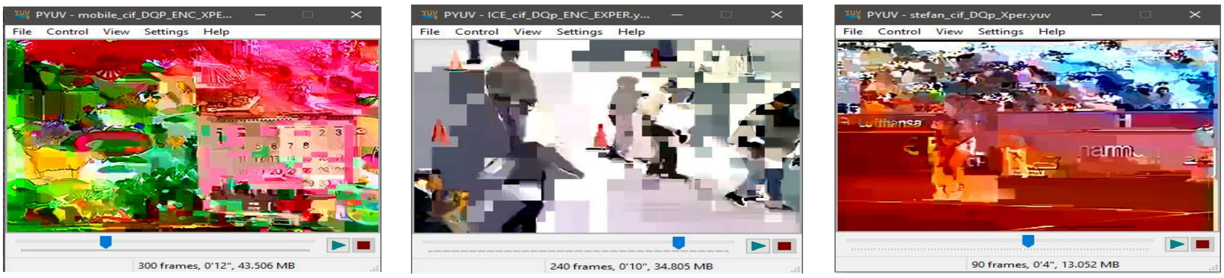
(a1)Original MOBILE Video (b1) Original ICE Video (c1) Original STEFANVideo

(a2) (b2) (c2) SE on only Motion with EXPer

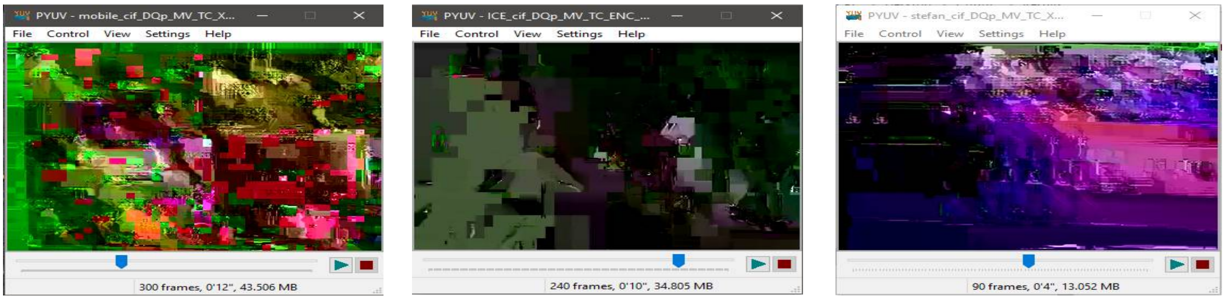
(a3) (b3) (c3) SE on only Texture with EXPer

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at Sensors 2019, 19, 1228; doi:10.3390/s19051228



(a4) (b4) (c4) SE on only delta QP with EXPer



(a5) (b5) (c5) SE on Combined (Motion, Texture and Delta QP) parameters with EXPer

Figure 5. Visual protection of tested videos with EXPer encryption on selected syntax elements

4.3. Computational Cost Analysis

To analyze the performance of the EXPer, the absolute encryption time in seconds was measured for CIF videos. Table 5 shows that the time is negligible compared to the compression time. Thus, the results show that EXPer encrypts the videos with a low computational cost, which is on average 3.1 % of the H.264/AVC compression time (when encoded with CABAC) without encryption. Notice that the absolute encryption time is taken separately to the Encoding time (Table 5, col 4)

$$\textit{Computational Cost} = \textit{Encoding Time} + \textit{Absolute Encryption Time} \qquad (9)$$

Table 5. Absolute encryption time (s) with H.264/AVC CABAC entropy coding on selected parameters (motion, texture, delta QP and their combination) with EXPer.

Sr. #	Videos (CIF)	File size (MB)	Encoding time with H.264/AVC using CABAC (Compression without encryption)	Absolute encryption time by applying SE on CABAC (sec.)						
				Only Motion	Only Texture	Only Delta QP	Motion and Texture	Motion and Delta QP	Delta QP and Texture	Motion , Texture and Delta QP
1.	Bus	21.7	73.697 s	1.896	1.044	1.195	1.249	0.342	1.433	2.431
2.	Coastguard	43.5	133.828 s	1.622	0.177	4.045	3.655	3.831	3.528	4.012
3.	Container	43.5	116.386 s	2.687	2.691	2.656	3.217	3.058	2.666	3.947
4.	Crew	43.5	156.804 s	1.601	1.803	4.151	2.928	3.373	3.079	4.512
5.	Flower	36.3	107.977 s	1.849	2.745	1.059	2.908	0.143	3.561	3.772
6.	Football	37.7	143.975 s	1.154	1.149	2.995	1.866	0.129	0.989	2.958
7.	Foreman	43.5	134.769 s	2.419	2.285	2.793	0.993	5.483	3.603	5.988
8.	Hall	36.3	118.614 s	1.198	3.267	3.554	3.141	3.611	3.25	3.984
9.	ICE	34.8	105.232 s	0.947	2.694	3.54	3.4	2.998	2.625	3.229
10.	Mobile	43.5	129.394 s	1.786	2.773	2.809	0.893	3.683	2.038	3.186
	Average	**absolute encryption time**		**1.715**	**2.062**	**2.879**	**2.226**	**2.639**	**2.677**	**3.801**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

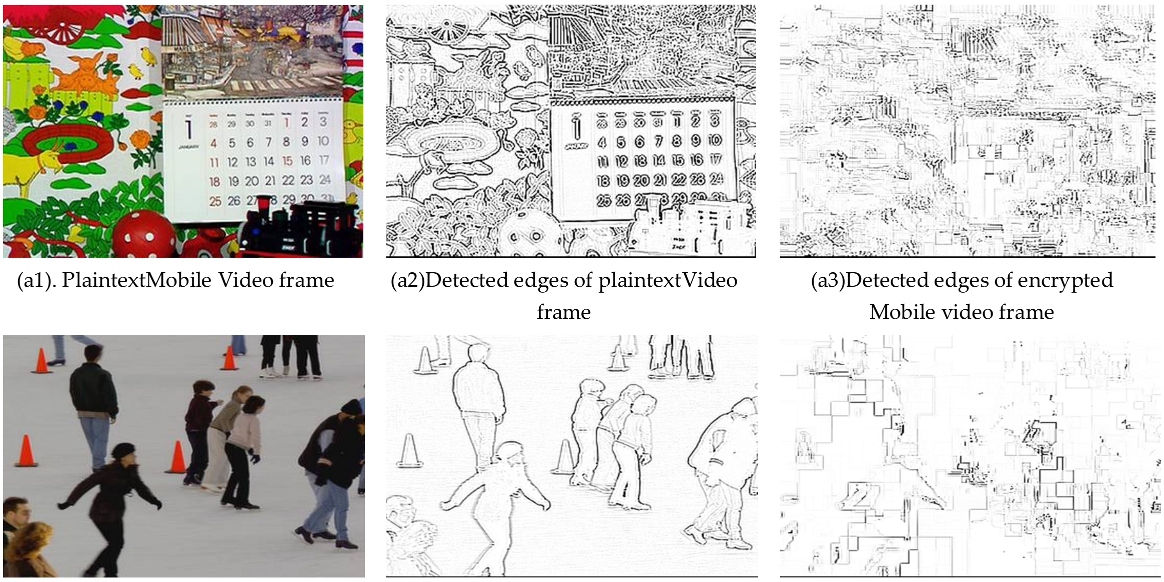
Peer-reviewed version available at Sensors 2019, 19, 1228; doi:10.3390/s19051228

4.4. Security Analysis of EXPer

The results obtained from various experiments in this Section validate the robustness of EXPer.

4.4.1. Perceptual Security

Perceptual quality is considered an important check on the strength of encryption algorithms. An encryption algorithm is considered robust if it succeeds in distorting a video sequence in such a way that an observer visually fails to detect any useful information from the encrypted bitstream. Clearly, the term 'useful' is dependent on the purpose that the video is to be put to, which, herein, is assumed to be IoMT purposes. The visual results of Figure. 5 already show that the video sequences encrypted with EXPer produce distorted results compared to the original video sequence. Furthermore, to evaluate the structural distortion of the proposed algorithm, 3 × 3 Laplacian edge detection [57] was performed. The detected edges of the plaintext video and encrypted video frames are illustrated in Figure. 6. The comparative results in Figure.6.a2, b2 with those of Figure. 6.a3, b3 show that the SE with EXPer distorts the video in a way that the attacker cannot easily acquire similarity information from edges of the encrypted video.



(a1). PlaintextMobile Video frame | (a2)Detected edges of plaintextVideo frame | (a3)Detected edges of encrypted Mobile video frame

(b1). Plaintext ICE video frame | (b2). Detected edges of plaintext ICE video frame | (b3). Detected edges of encrypted ICE video frame

Figure. 6. The comparative visual impact of EXPer on ICE and Mobile video sequences after Laplacian edge detector

Peak Signal to Noise Ratio (PSNR)

PSNR [58] measures the maximum possible absolute differences between the original bitstream and the encrypted bitstream in decibels and is calculated as:

$$\text{PSNR} = 10.\log_{10}\frac{\text{MaxErr}^2}{\sum_{i=1,j=1}^{m,n}(X_{i,y} - Y_{i,j})2} \qquad (10)$$

where 'm' and 'n' are the width and height of the video frame under consideration, while 'X' and 'Y' represent the pixel's intensity values of the two frames being compared. For a video sequence, the PSNRs of the frames are averaged across the sequence. A lower value of PSNR indicates less similarity between an original video sequence and the video sequence reconstructed from a compressed and encrypted bitstream.

Table 6 (col. 2, 3) demonstrates the average (arithmetic mean) PSNR of test video sequences after SE (with EXPer) and without encryption (video only compressed). The results show that the average

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 1228; doi:10.3390/s19051228

PSNR value is much lower. Hence, the proposed SE with EXPer encryption produced the highly distorted video. Thus, EXPer can be considered for video protection in IoMT.

Pixel-correlation analysis

Another statistical method to compute the similarity between the original and encrypted pixels of the video frame is cross-correlation. The cross-correlation coefficient, r, is calculated as:

$$r = \frac{\sum_m \sum_n (X_{mn}-\bar{X})(Y_{mn}-\bar{Y})}{\sqrt{(\sum_m \sum_n (X_{mn}-\bar{X})^2)(\sum_m \sum_n (Y_{mn}-\bar{Y})^2)}} \quad (11)$$

where $\bar{X}$, $\bar{Y}$ are the mean intensity values of pixels the original and distorted video frames. The values of the r ranges from 1 to -1. When two frames are the same, the correlation index is at a maximum, which is 1. Therefore, a lower value of correlation coefficient indicates higher distortion as a result of encryption. Table 6 (cols. 4 and 5) presents calculated cross correlations between encrypted and compressed video frames. The average value of the pixel correlations among the plaintext and encrypted video frames is near to zero, confirming that video sequences encrypted with EXPer are considerably distorted in a statistical sense, thus providing good confidentiality.

Table 6. Comparison of average PSNR (dB) and pixel cross-correlation of SE at QP= 36 with EXPer for the videos encoded with CABAC

Video sequence	Average PSNR		Average pixel cross-correlation	
	Compressed	Encrypted	Compressed	Encrypted
Football	[Y:31.63,U:37.45 V: 39.01]	[Y:10.56,U:18.97,V:19.25]	0.9816	0.04708
ICE	[Y:29.18,U:34.16,V: 33.47]	[Y:6.27,U: 14.99,V: 11.64]	0.9991	-0.0173
Mobile	[Y:32.86 ,U:38.06,V:39.47]	[Y:6.00,U: 21.97,V: 27.32]	0.9923	0.014

The correlation between adjacent pixels within video frames in the different directions (horizontal, vertical and diagonal) for plaintext and encrypted Mobile video are shown in Figure 7. The correlation test is performed by taking randomly N = 6000 pairs of adjacent pixels from the original and selectively encrypted test video frames.

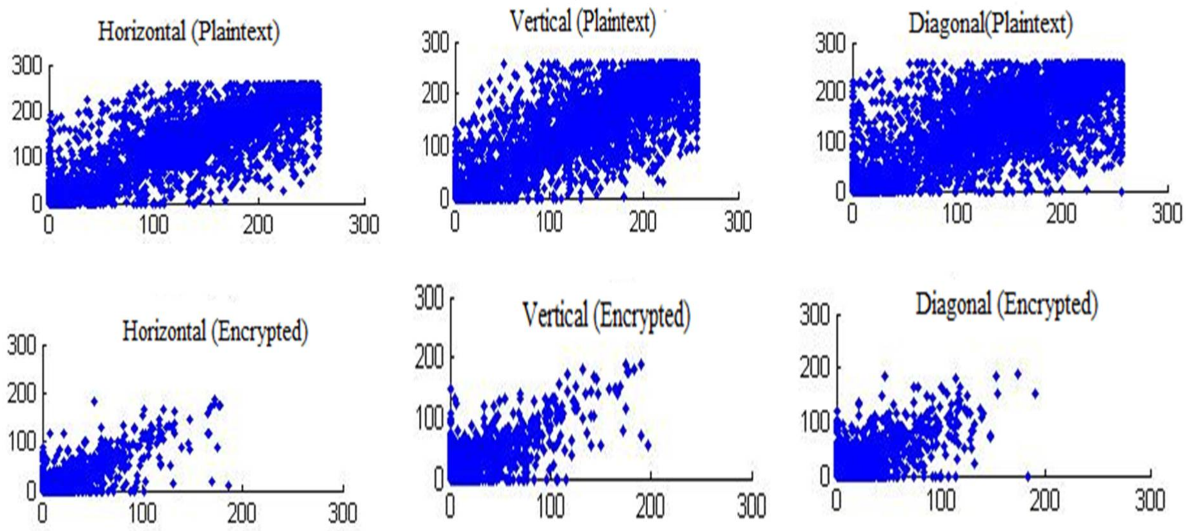


Figure 7. Pixels correlation (horizontal, vertical and diagonal) for the plaintext and encrypted Mobile Video with EXPer

Structural Similarity Index (SSIM)

The SSIM index [58] is a metric which gauges the structural similarity between original and reconstructed video frames, having a range normally from 0 to1. Values of SSIM nearer to 0 means

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at Sensors 2019, 19, 1228; doi:10.3390/s19051228

less structural similarity between the plaintext and the reconstructed encrypted bitstream, which means greater distortion has occurred. Values nearer to 1 means more structural similarity. The SSIM values on videos by applying SE with EXPer are reported in Figure. 8. The SSIM plots make clear that videos are drastically changed when EXPer is applied on selected combined parameters and that it would be extremely difficult to extrapolate the encrypted parts.

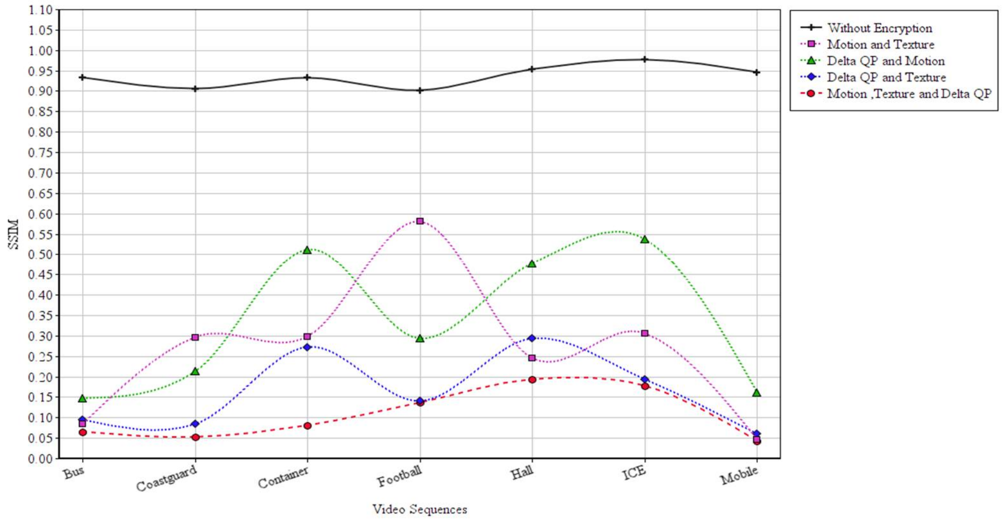


Figure. 8. Average SSIM values for EXPer on different combinations of motion, texture and deltaQP parameters on test videos

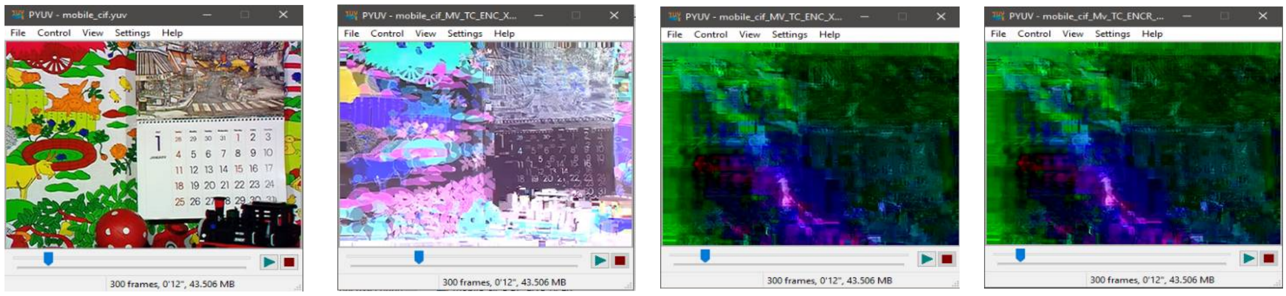
5. Comparison of EXPer with state-of-art ciphers

To confirm the worth of the EXPer, a comparison with the most commonly used encryption algorithms XOR and AES was performed. XOR can be considered the most suitable encryption algorithm for IoT applications due to its simplicity and lower computational complexity. However, XOR provides limited confidentiality for images, due to potentially high cross-correlations, and therefore AES can be utilized to provide greater confidentiality. Though AES is robust against known-plaintext, brute force, and statistical attacks, it incurs a higher encoding and decoding overhead, which is expensive for resource-constrained IoMT devices.

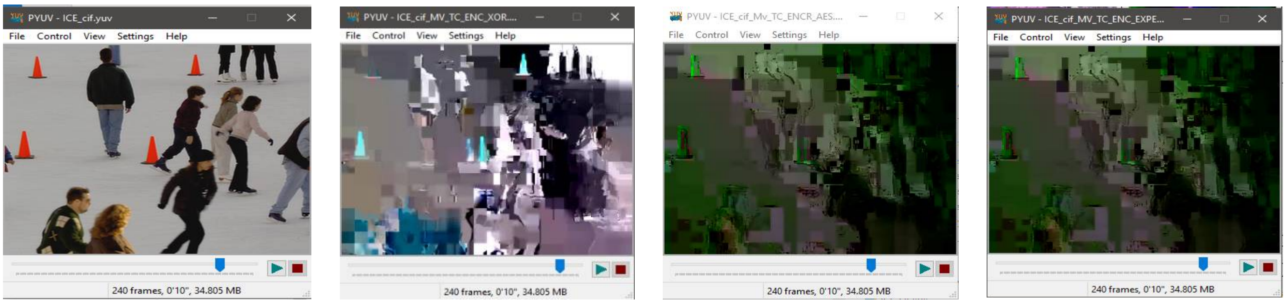
The results were taken with both state-of-art ciphers to compare their performance with EXPer. Comparative visual results with XOR, AES-CFB [11], and EXPer with CABAC coding are presented in Figure 9. Figure 9 (b1–b3), (c1 – c3) and (d1 – d3) depict SE of the videos with three ciphers. The comparative results in Figure 9 (c vs. d) imply that the EXPer provides the same level of visual protection and robustness as AES. It is worth mentioning here that dQP encryption is not applied in these comparative results as AES rounds make the dQP encrypted video non-format compliant.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 1228; doi:10.3390/s19051228



(a1) Original Raw Mobile Video (b1). Encryption with XOR (c1). Encryption with AES-CFB (d1). Encryption with EXPer



(a2) Original Raw ICE Video (b2). Encryption with XOR (c2). Encryption with AES-CFB (d2).Encryption with EXPer

(a3) Original Raw Football Video (b3). Encryption with XOR (c3). Encryption with AES-CFB (d3).Encryption with EXPer

Figure 9. Comparative visual effects of selective encryption (on I, P and B frames) on combined motion + texture syntax elements with XOR, AES-CFB and EXPer on test videos at QP= 36 encoded with CABAC

5.1. Comparative Visual Quality Analysis

For video quality analysis of these three ciphers, PSNR and SSIM results were also taken. A PSNR comparison among XOR, AES-CFB, and EXPer with combined motion and texture parameters encryption on different QPs is given in Table 7. The luminance (Y)-PSNR of the Mobile video sequence is 7.18 dB, 6.07 dB, and 6.27dBwith XOR, AES and EXPer respectively. While a noticeable point here is that EXPer is able to encrypt additional syntax element in implemented SE (Section 4), so the Y-PSNR of EXPer is 6.00 B (Table 7 (row 4, col. 3), less than AES-CFB, which is 6.07 dB for the Mobile video. The comparative PSNR results confirm that the proposed algorithm produces PSNR values almost equivalent to AES.

The SSIM of the encrypted video with combined motion and texture parameters is illustrated in Figure 9. The comparative results show that video sequence encrypted with EXPer and AES-CFB has smaller SSIM values than encryption with XOR. Lower SSIM values indicate more content protection. Furthermore, from the evidence of Figure 10, EXPer provides almost the same level of confidentiality as AES-CFB.

Hence, the PSNR and SSIM results imply that the encryption applied with EXPer provides confidentiality similar to the AES.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 1228; doi:10.3390/s19051228

Table 7. Comparison of average PSNR (dB) of SE with XOR, AES-CFB and EXPer at four different QP levels

Videos	QP	Encoded Without SE			SE with XOR			SE with AES-CFB			SE with EXPer		
		Y	U	V	Y	U	V	Y	U	V	Y	U	V
Football	12	43.9	46.71	47.65	9.72	13.08	20.92	9.52	18.25	21.13	9.53	19.25	22.13
	24	37.03	41.55	42.76	8.67	13.28	20.75	10.04	18.06	19.36	10.16	19.06	20.36
	36	31.63	37.45	39.01	9.94	13.19	20.87	10.36	17.97	18.24	10.56	18.97	19.25
	48	28.72	34.99	37.1	10.63	12.94	21.07	11.31	17.38	16.9	11.51	18.4	17.96
Mobile	12	42.72	45.18	44.96	7.05	13.26	13.95	6.06	12.47	10.46	6.26	13.47	11.46
	24	34.69	38.46	38.03	7.09	13.2	14.04	6.07	13.18	10.79	6.19	14.18	11.79
	36	29.18	34.16	33.47	7.18	13.09	14.16	6.07	13.98	10.62	6.27	14.99	11.64
	48	24.96	31.89	31.01	7.54	14.83	13.97	7.05	15.3	11.83	7.08	16.35	12.84
ICE	12	44.74	48.77	49.56	7.91	18.82	25.85	5.74	20.39	26.43	5.94	21.39	27.43
	24	38.78	43.21	44.02	7.73	18.33	26.64	5.8	20.97	26.32	6	21.97	27.32
	36	32.86	38.06	39.47	7.93	18.74	27	6.01	21.98	23.75	6.07	22.98	24.77
	48	29.09	36.03	37.58	8.25	18.21	27.66	5.9	18.54	19.25	5.98	19.57	20.27

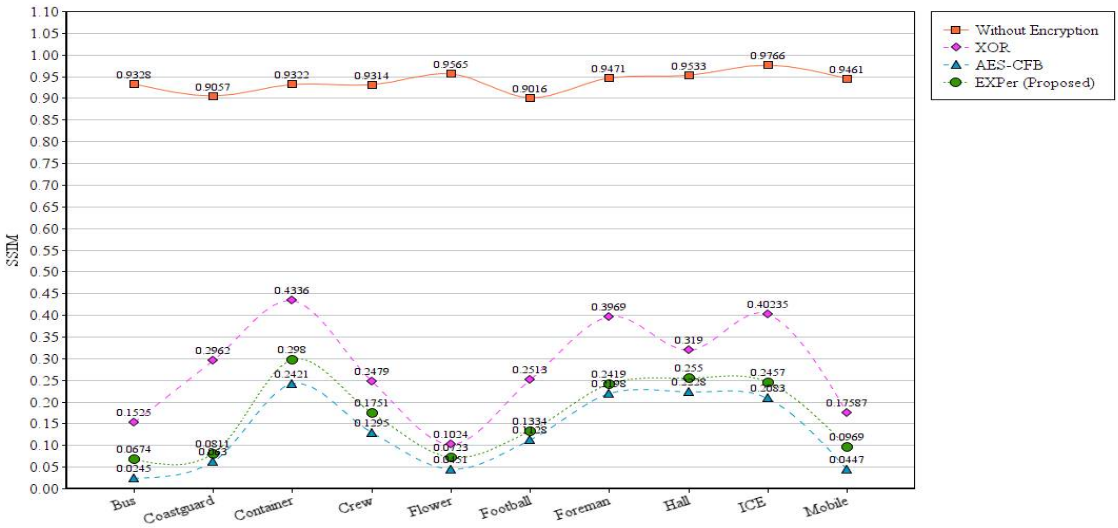


Figure 10. Comparison between SSIM values for XOR, AES-CFB and EXPer for selective encryption with combined motion and texture parameters at QP= 36 on 10 tested videos

5.2. Comparative Computational Efficiency

In addition to the visual content protection, the efficiency of an encryption algorithm for real-time processing is dependent on the execution/encoding time. Therefore, to evaluate the efficiency of EXPer, a comparison with the encoding time of standard algorithms XOR and AES-CFB have been performed. The comparative results of Figure 10 show that the absolute encoding time for only motion parameters encryption and only texture parameters encryption with EXPer is 91.35 s and 87.34 s respectively, which is less than AES-CFB for the ICE video. Likewise, the encoding time for combined motion and texture parameters encryption is 89.41 s, lower than AES-CFB. The graphical results of Figure 11 indicate that the absolute encoding time with EXPer encryption is nearly equivalent to encryption with XOR. Thus, the efficiency of EXPer in terms of execution time is distinctly better than AES-CFB. EXPer provides an almost similar level of protection to that provided by AES-CFB but has a very small computational overhead.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at Sensors 2019, 19, 1228; doi:10.3390/s19051228

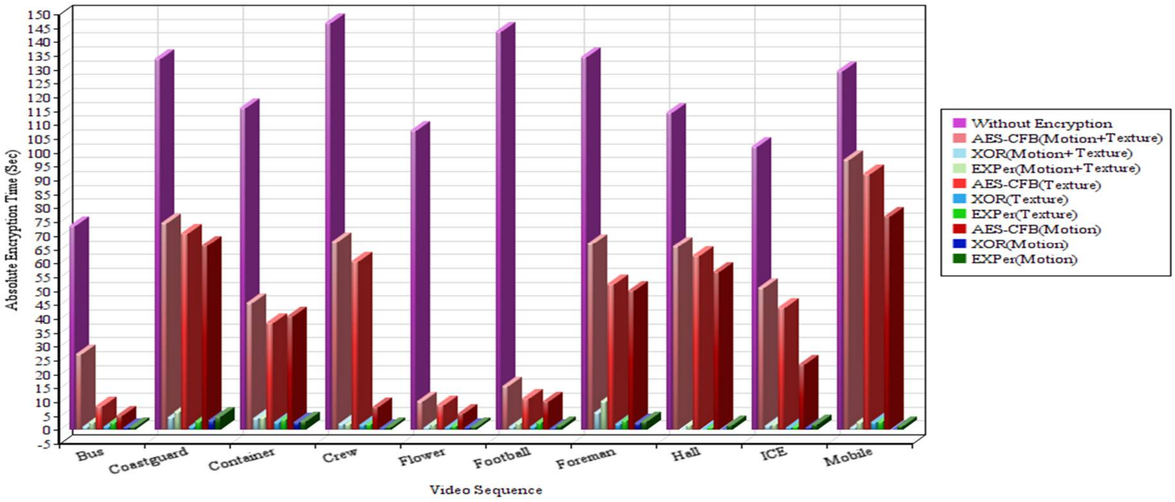
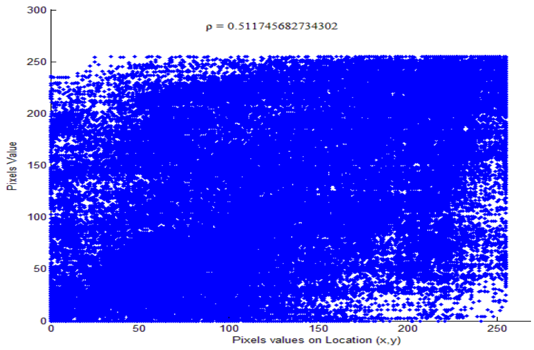


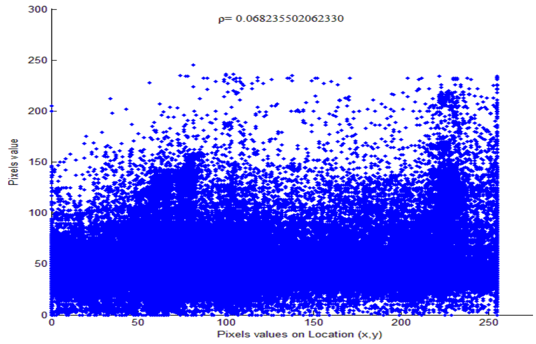
Figure 11. Comparative absolute encoding time of EXPer with AES-CFB and XOR

5.3. Comparative Security Analysis

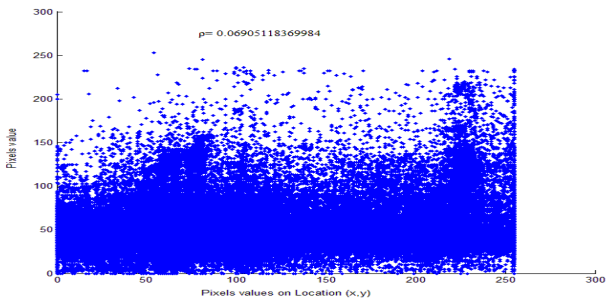
For the security analysis of EXPer, a comparison with the correlation coefficient of standard algorithms XOR and AES-CFB has been performed. Figure. 12 shows comparative correlation coefficients of plaintext Mobile video frame and encrypted mobile video frame. The results show that frame encrypted with the EXPer has pixels correlation coefficient value ρ= 0.06905118369984 which is almost equivalent to the pixels correlation value of the frame encrypted with AES-CFB that is 0.068235502062. This implies that EXPer has achieved the same level of randomness as AES-CFB and outperform the XOR with correlation value ρ=0.511745682734. This demonstrates that the proposed EXPer has greater potential to resist against the statistical attacks.



(a) XOR



(b) AES-CFB



(c) EXPer

Figure 12. Comparative pixels correlation coefficient of plaintext and encrypted Mobile video

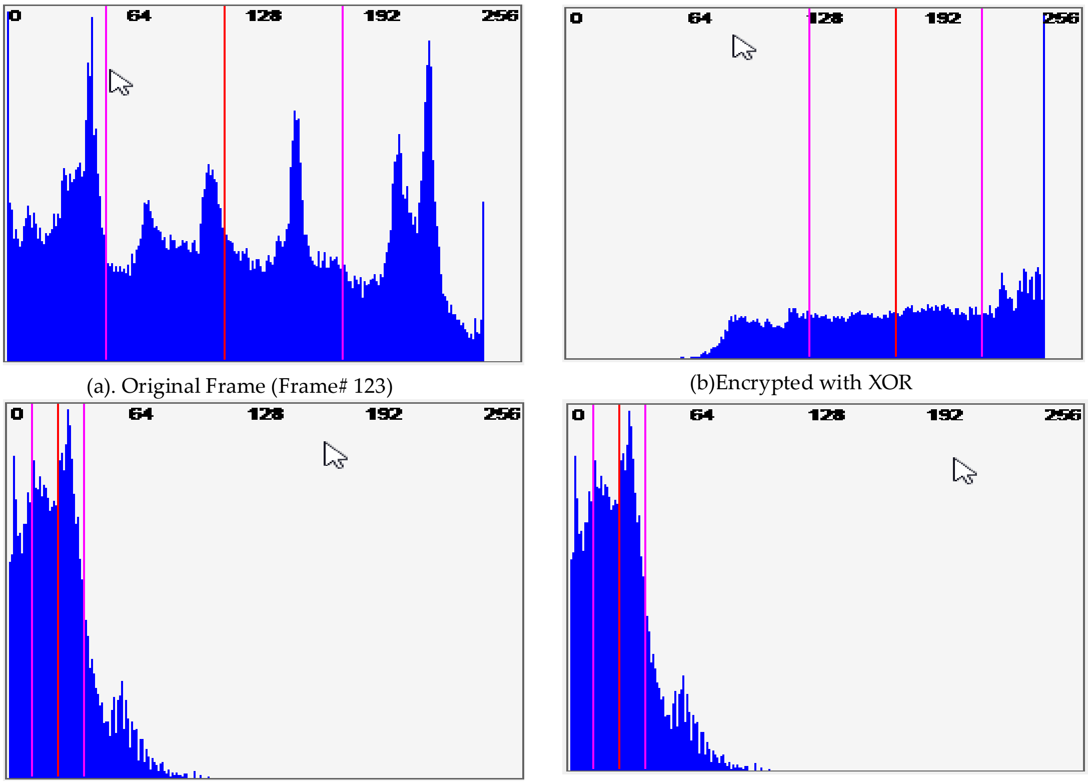
Peer-reviewed version available at Sensors 2019, 19, 1228; doi:10.3390/s19051228

5.4. Comparative Entropy analysis

Entropy defines uncertainty or the chaos level within the video frames. It measures the amount of the grey level and the probability corresponding to the total information inside all other pixels and determines which pixels carry most of the information. It is calculated as:

$$H(f) = -\sum_{i=0}^{2N-1} p(f) \log_2(f) \tag{12}$$

where 'f' is the amount of gray level and 'p(f)' is the probability of 'f'. Figure 13 illustrate comparative entropy histogram of Mobile video frames encrypted with XOR, AES and EXPer. This entropy is evident in the spreading of more black and sharp colours (shown in Figure 5 and 9) across the video frames compared to the original histogram (Figure 13 (a)) values prior to applied SE with three ciphers over Mobile Video. This finding implies that, if the videos are selectively encrypted by EXPer and AES, it is difficult to infer the presence of an object in any one of the R, G, B and luminance domains. The results also illustrate that the EXper has attained the security against inference attacks equivalent to the AES-CFB.



(a). Original Frame (Frame# 123) (b)Encrypted with XOR

(b)Encrypted with AES (b)Encrypted with EXPer (Proposed)

Figure 13. Comparative entropy of plaintext and encrypted Mobile video

6. Conclusion

Various classes of IoMT devices are utilized for multiple services i.e. stored video streaming (YouTube, online lectures), live video streaming in the cases of video conferencing, online gaming etc., with the most confidential being real-time interactive video streaming in the form of surveillance applications. It is crucial to modern IoMT nodes to provide data confidentiality in the form of data encryption. The most reliable cipher is AES with 128/192/256 bit keys. However, AES is still not an optimal choice for low powered surveillance devices with simple hardware. In this paper, by keeping in view current security needs, we propose an ESR-validated security scheme for IoMT devices. Within a security scheme, the contribution of the paper is to examine two alternative entropy coders available for the H.264 codec, i.e. CAVLC and CABAC in detail and determine the ESR when applying SE with the proposed cipher to encrypt the selected syntax elements. As identified herein, using the CABAC entropy coder (see Table 1) can considerably save on the ESR percentage that is the maximum is 14.61%, while the CAVLC is 29.27%, so that the equivalent average ESR is 12% for CAVLC and 7.5% for CABAC across the tested reference videos. The ESR calculated for CABAC is acceptable for IoMT applications as reduced encryption data consumes less computation during encoding. In the proposed security scheme, a novel cipher, EXPer, works on cryptographic basic principles i.e. permutation and XOR with three different 128-bit keys over selected syntax elements of CABAC encoder. EXPer even performs very well on absolute values of syntax elements i.e. dQP, without changing the bit-rate and crashing the decoder (if the decoder is applied to encrypted video). Comparative analysis with the existing state-of-art ciphers shows that EXPer yields confidentiality almost similar to that of AES-CFB, but the computational cost is similar to the XOR, which makes it a suitable choice for protecting real-time video communication in an IoMT setting. Our detailed security analysis revealed that the proposed EXPer is robust enough against multiple attacks.

Future work based on taking measurements of ESR in this paper can provide a way to more precisely model the trade-offs between computational complexity, and memory access in terms of energy consumption within a video sensor device. Both entropy coders have a content dependency, which increases the effect of bit errors. This implies that error resiliency or channel coding should be built into transmission over an IoMT network along with proper key management solution in future.

Acknowledgment: This research paper is produced as part of a government-funded project (National Research Program for Universities (NRPU-2016)) with no: 6282/Punjab/NRPU/R&D/HEC/2016. We appreciate the support of the Higher Education Commission (HEC) of Pakistan for this project.

Author Contributions: Conceptualization, Mamoona Asghar; Formal analysis, Amna Shifa and Salma Noor; Funding acquisition, Mamoona Asghar; Investigation, Salma Noor; Methodology, Amna Shifa and Martin Fleury; Software, Amna Shifa; Supervision, Mamoona Asghar; Validation, Neelam Gohar; Visualization, Amna Shifa and Neelam Gohar; Writing – original draft, Amna Shifa; Writing – review & editing, Mamoona Asghar and Martin Fleury.

References

1. Al-Fuqaha, A.; Guizani, M.; Mohammadi, M.; Aledhari, M.; Ayyash, M. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Commun. Surv. Tutorials* **2015**, *17*, 2347–2376.
2. Alvi, S.A.; Afzal, B.; Shah, G.A.; Atzori, L.; Mahmood, W. Internet of multimedia things: Vision and challenges. *Ad Hoc Networks* 2015.
3. Aslam, A.; Curry, E. Towards a Generalized Approach for Deep Neural Network Based Event Processing for the Internet of Multimedia Things. *IEEE Access* **2018**.
4. Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. *Futur. Gener. Comput. Syst.* **2013**, *29*, 1645–1660.
5. Sicari, S.; Rizzardi, A.; Grieco, L.A.; Coen-Porisini, A. Security, privacy and trust in Internet of Things: The road ahead. *Comput. Networks* **2015**, *76*, 146–164.
6. Beginning of a New Standard: Internet of Media Things. *KSII Trans. Internet Inf. Syst.* **2017**, *11*.
7. Shifa, A.; Asghar, M.N.; Fleury, M. Multimedia security perspectives in IoT. In Proceedings of the 2016 Sixth International Conference on Innovative Computing Technology (INTECH); IEEE, 2016; pp. 550–555.
8. Semertzidis, T.; Dimitropoulos, K.; Koutsia, A.; Grammalidis, N. Video sensor network for real-time traffic monitoring and surveillance. *IET Intell. Transp. Syst.* **2010**, *4*, 103.
9. Asghar, M.N.; Kousar, R.; Majid, H.; Fleury, M. Transparent encryption with scalable video communication: Lower-latency, CABAC-based schemes. *J. Vis. Commun. Image Represent.* **2017**, *45*, 122–136.
10. Shifa, A.; Afgan, M.S.; Asghar, M.N.; Fleury, M.; Memon, I.; Abdullah, S.; Rasheed, N. Joint Crypto-Stego Scheme for Enhanced Image Protection With Nearest-Centroid Clustering. *IEEE Access* **2018**, *6*, 16189–16206.
11. Technology, N.I. of S. and FIPS-197: Advanced Encryption Standard (AES). *Fed. Inf. Process. Stand. Publ.* 2001.
12. Long, M.; Peng, F.; Li, H. Separable reversible data hiding and encryption for HEVC video. *J. Real-Time Image Process.* **2018**, *14*, 171–182.

13. Shahid, Z.; Puech, W. Visual Protection of HEVC Video by Selective Encryption of CABAC Binstrings. *IEEE Trans. Multimed.* **2014**, *16*, 24–36.
14. Adeyemi-Ejeye, A.O.; Alreshoodi, M.; Al-Jobouri, L.; Fleury, M. Prospects for live higher resolution video streaming to mobile devices: achievable quality across wireless links. *J. Real-Time Image Process.* **2018**, 1–15.
15. Lookabaugh, T.; Sicker, D.C. Selective encryption for consumer applications. *IEEE Commun. Mag.* **2004**, *42*, 124–129.
16. Asghar, M.N.; Fleury, M.; Makki, S. Interoperable conditional access with video selective encryption for portable devices. *Multimed. Tools Appl.* **2017**, *76*, 13139–13152.
17. Dufaux, F. Video Scrambling for Privacy Protection in Video Surveillance - Recent Results and Validation Framework. *SPIE Defense, Secur. Sens.* **2011**.
18. Fei Peng; Xiao-wen Zhu; Min Long An ROI Privacy Protection Scheme for H.264 Video Based on FMO and Chaos. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 1688–1699.
19. Massoudi, A.; Lefebvre, F.; De Vleeschouwer, C.; Macq, B.; Quisquater, J.-J. Overview on Selective Encryption of Image and Video: Challenges and Perspectives. *EURASIP J. Inf. Secur.* **2008**, *2008*, 1–18.
20. Zhang, Z.-K.; Cho, M.C.Y.; Wang, C.-W.; Hsu, C.-W.; Chen, C.-K.; Shieh, S. IoT Security: Ongoing Challenges and Research Opportunities. In Proceedings of the 2014 IEEE 7th International Conference on Service-Oriented Computing and Applications; IEEE, 2014; pp. 230–234.
21. M. Ghanbari Standard Codecs: Image Compression to Advanced Video Coding ; IET 2003. London, UK.; Available online: https://dl.acm.org/citation.cfm?id=1237948 (accessed on Dec 22, 2018).
22. Wiegand, T.; Sullivan, G.J.; Bjontegaard, G.; Luthra, A. Overview of the H.264/AVC video coding standard. *IEEE Trans. Circuits Syst. Video Technol.* **2003**, *13*, 560–576.
23. Stutz, T.; Uhl, A. A Survey of H.264 AVC/SVC Encryption. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 325–339.
24. 24. Schwarz, H.; Marpe, D.; Wiegand, T. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *IEEE Trans. Circuits Syst. Video Technol.* 2007, *17*, 1103–1120
25. Shahid, Z.; Chaumont, M.; Puech, W. Fast Protection of H.264/AVC by Selective Encryption of CAVLC and CABAC for I and P Frames. *IEEE Trans. Circuits Syst. Video Technol.* **2011**, *21*, 565–576.
26. 26. Bjøntegaard, G. ; Lillevold, K. Context-adaptive VLC coding of coefficients.**2002**, JVT document JVT-C028.
27. Marpe, D.; Schwarz, H.; Wiegand, T. Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard. *IEEE Trans. Circuits Syst. Video Technol.* **2003**, *13*, 620–636.
28. Sullivan, G.J.; Ohm, J.-R.; Han, W.-J.; Wiegand, T. Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 1649–1668.
29. Shahid, Z.; Puech, W. Visual Protection of HEVC Video by Selective Encryption of CABAC Binstrings. *IEEE Trans. Multimed.* **2014**, *16*, 24–36.
30. Usman, M.; Jan, M.A.; He, X.; Nanda, P. Data Sharing in Secure Multimedia Wireless Sensor Networks. In Proceedings of the 2016 IEEE Trustcom/BigDataSE/ISPA; IEEE, **2016**; pp. 590–597.
31. Raspberry Pi: https://projects.raspberrypi.org/en/projects/getting-started-with-picamera/7 (Last Accessed on 30 Nov. 2018)
32. Li S; Chen, G.; Zheng, X .Chaos-based encryption for digital images and videos. In Furht, B and Kirovski, D (Eds.).Multimedia Security Handbook ; **2005;** Boca Raton, FL: CRC Press, pp. 133-168
33. Liu, F.; Koenig, H. A survey of video encryption algorithms. *Comput. Secur.* **2010**, *29*, 3–15.
34. El Assad, S.; Farajallah, M. A new chaos-based image encryption system. *Signal Process. Image Commun.* **2016**, *41*, 144–157.
35. Salami, S. Al; Baek, J.; Salah, K.; Damiani, E. Lightweight Encryption for Smart Home. In Proceedings of the 2016 11th International Conference on Availability, Reliability and Security (ARES); IEEE, 2016; pp. 382–388.
36. Yao, X.; Chen, Z.; Tian, Y. A lightweight attribute-based encryption scheme for the Internet of Things. *Futur. Gener. Comput. Syst.* **2015**, *49*, 104–112.
37. Xin, M. A Mixed Encryption Algorithm Used in Internet of Things Security Transmission System. In Proceedings of the 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery; IEEE, 2015; pp. 62–65.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0308.v1

Peer-reviewed version available at Sensors 2019, 19, 1228; doi:10.3390/s19051228

38. Prasetyo, K.N.; Purwanto, Y.; Darlis, D. An implementation of data encryption for Internet of Things using blowfish algorithm on FPGA. In Proceedings of the 2014 2nd International Conference on Information and Communication Technology (ICoICT); IEEE, 2014; pp. 75–79.
39. Lee, J.-Y.; Lin, W.-C.; Huang, Y.-H. A lightweight authentication protocol for Internet of Things. In Proceedings of the 2014 International Symposium on Next-Generation Electronics (ISNE); IEEE, 2014; pp. 1–2.
40. Mahalle, P.N.; Anggorojati, B.; Prasad, N.R.; Prasad, R.R. Identity Authentication and Capability Based Access Control (IACAC) for the Internet of Things. *J. Cyber Secur. Mobil.* **2013**, *1*, 309–348
41. Vinayaga Sundaram, B.; Ramnath M.; Prasanth M.; Varsha Sundaram J. Encryption and hash based security in Internet of Things. In Proceedings of the 2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN); IEEE, 2015; pp. 1–6.
42. Bogdanov, A.; Knudsen, L.R.; Leander, G.; Paar, C.; Poschmann, A.; Robshaw, M.J.B.; Seurin, Y.; Vikkelsoe, C. PRESENT: An Ultra-Lightweight Block Cipher. In Cryptographic Hardware and Embedded Systems - CHES 2007; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; pp. 450–466.
43. Engels, D.; Saarinen, M.-J.O.; Schweitzer, P.; Smith, E.M. The Hummingbird-2 Lightweight Authenticated Encryption Algorithm. In; Springer, Berlin, Heidelberg, 2012; pp. 19–31.
44. Suzaki, T.; Minematsu, K.; Morioka, S.; Kobayashi, E. $\textnormal{\textsc{TWINE}}$: A Lightweight Block Cipher for Multiple Platforms. In; Springer, Berlin, Heidelberg, 2013; pp. 339–354.
45. Albrecht, M.R.; Driessen, B.; Kavun, E.B.; Leander, G.; Paar, C.; Yalçın, T. Block Ciphers – Focus on the Linear Layer (feat. PRIDE). In; Springer, Berlin, Heidelberg, 2014; pp. 57–76.
46. Janakiraman, S.; Thenmozhi, K.; Rayappan, J.B.B.; Amirtharajan, R. Lightweight chaotic image encryption algorithm for real-time embedded system: Implementation and analysis on 32-bit microcontroller. Microprocess. Microsyst. 2018, 56, 1–12.
47. Noura, H.; Chehab, A.; Sleem, L.; Noura, M.; Couturier, R.; Mansour, M.M. One round cipher algorithm for multimedia IoT devices. Multimed. Tools Appl. 2018, 77, 18383–18413.
48. Lei, B.Y.; Lo, K.T.; Haijun Lei A new H.264 video encryption scheme based on chaotic cipher. In Proceedings of the 2010 International Conference on Communications, Circuits and Systems (ICCCAS); IEEE, 2010; pp. 373–377.
49. François, M.; Grosges, T.; Barchiesi, D.; Erra, R. A new image encryption scheme based on a chaotic function. *Signal Process. Image Commun.* **2012**, *27*, 249–259.
50. Bellovin, S.M.; Housley, R. Guidelines for Cryptographic Key Management.
51. Asghar, M.N.; Ghanbari, M.; Fleury, M.; Reed, M.J. Sufficient encryption based on entropy coding syntax elements of H.264/SVC. *Multimed. Tools Appl.* **2015**, *74*, 10215–10241.
52. Asghar, M.N.; Ghanbari, M. An Efficient Security System for CABAC Bin-Strings of H.264/SVC. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 425–437.
53. Khan, J.S.; Ahmad, J. Chaos based efficient selective image encryption. Multidimens. Syst. Signal Process. 2018, 1–19.
54. Asghar, M.N.; Ghanbari, M.; Fleury, M.; Reed, M.J. Confidentiality of a selectively encrypted H.264 coded video bit-stream. J. Vis. Commun. Image Represent. 2014, 25, 487–498.
55. https://media.xiph.org/video/derf/(Last Accessed on 30 Nov. 2018)
56. 54.https://www.hhi.fraunhofer.de/en/departments/vca/research-groups/image-video-coding/research-topics/svc-extension-of-h264avc/jsvm-reference-software.html (Last Accessed on 30 Nov. 2018)
57. Bhardwaj, S.; Mittal, A. A Survey on Various Edge Detector Techniques. *Procedia Technol.* **2012**, *4*, 220–226.
58. Huynh-Thu, Q.; Ghanbari, M. The accuracy of PSNR in predicting video quality for different video scenes and frame rates. *Telecommun. Syst.* **2012**, *49*, 35–48.
59. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612.