Indian Journal of Science and Technology, *Vol 8(10), 940–949, May 2015*

ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Combining Audio Samples and Image Frames for Enhancing Video Security

M. Al-Hazaimeh Obaida*

Department of Computer Science, Al-Balqa' Applied University, Jordan; dr_obaida@bau.edu.jo

Abstract

Advances in the digital content transmission have increased exponentially in the past few years. In addition, the video encryption plays an important role in guaranteeing data transmission security, this role is increasing continuously due to its involvement with the development of multimedia technology. In this paper, we propose applicable and secure video encryption algorithm. The proposed algorithm is mainly based on the symmetric key cryptographic technique, and it employs the audio sample to eliminate the need for trusted third party for key exchange process. The new algorithm starts working by splitting the undertaken video into audio samples and video frames. Next, the audio sample is used to generate the public key. The yielded audio sample and video frame are then encrypted along with the generated public key to produce the cipher data. The public key is then added to the cipher data randomly. The receiver on other side will extract firstly the public key from cipher data, and then complete the decryption process. Based on the results, and comparing with the existing encryption algorithms, the presented algorithm is able to maximize the tolerance skew value to be close to the optimal value. Moreover, the results showed that the proposed algorithm is efficient, and give high performance. The performance is measured by means of parameters such as Peak Signal to Noise Ratio (PSNR) and Mean Squared Error (MSE). The resulted MES values are close to zero, and the resulted values for the PSNR are between 30 and 58 dB, and those values proved the effectiveness of the proposed algorithm.

Keywords: Cryptography, Compression, Skew Tolerances, Video Encryption, Video Decryption

1. Introduction

The security of the multimedia systems is marked as a major issue in the networked infrastructure and the transmission of data through secluded systems. The security of the data is enhanced by use of cryptography, a technique that is used in protection of data that is generated via open networks. It also includes the transmission of unreadable data. The major drawback in this technique is that most of the algorithms used in the encryption and decryption only deal with text data. The latter makes the algorithms unsuitable for multimedia technologies including pay-TV, video conferencing, video broadcast, and Video-On Demand. Characteristically, the digital multimedia content is a combination of image frames, text, and audio elements[1,2].

In order to protect the contents of the digital multimedia, the security approaches that are software-based have presented viable solutions that can accommodate both read and write data access. This allows for data protection against unauthorized access. The encryption algorithms can either be symmetric or asymmetric. In this paper, the symmetric key encryption is implemented due to the sustainability of the technique with digital multimedia content encryption. Other encryption algorithms considered include the four categories of video encryption algorithms. These include: permutation encryption algorithm, completely encryption algorithms, and perceptional and selective encryption algorithms[3-6].

The completely encryption algorithm work by the initial compression of the entire video followed by encryption of the video data based on the traditional algorithms such as DES, AES, 3-DES, RC4, among others as shown in Figure 1. The computation process of the algorithm make them time consuming. This renders them unsuitable for the specific encryption needs[4].

*Author for correspondence

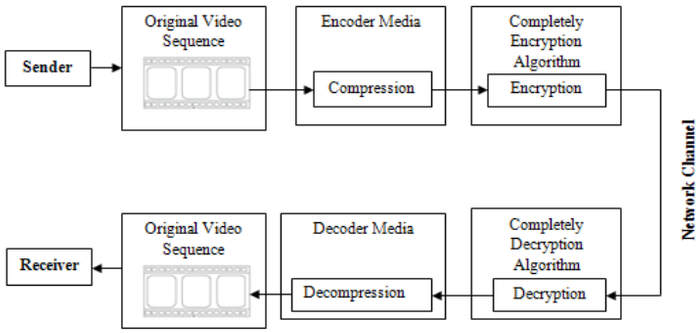


Figure 1. Structure of completely encryption algorithms.

Another encryption process discussed is the use of permutation. The process involves scrambling of the video content in audio, images, and text[7,8]. However, in the selective encryption algorithm, only the specific video byte that needs to be encrypted is considered. The algorithm works on the basis that some contents of the video are not critical. This leads to reduction of the encryption time as compared to the completely encryption algorithm techniques as shown in Figure 2. The other encryption algorithm covered in the project is the perpetual-based encryption algorithm. This enables the potential viewers to listen and view though with low quality of the video[6,9,10].

The encryption process concerning videos generally involves the use of more complicated technologies as compared to voice and text encryption techniques. The complication has made researchers in the subject to focus their attention more on video encryption than other forms of encryption. This includes the analysis of the obstacles in the video separation process. The separation process involves the separation of the video into discrete images[11]. Therefore, making the choice of the encryption technique to be used and adopted needs careful analysis. Many research studies have been conducted to assess the real-time streaming video elements for understanding of the desired speed for streaming.

Most of the encryption techniques are used for binary and text data for a number of reasons. Firstly, multimedia data, including video and image, are usually bulky and smaller in size. Encryption of bulk data by use of the traditional ciphers may incur significant overhead and may be too expensive for real-time applications such as image surveillance and video conferencing. Adjacent pixels in digital images also have similar values of gray scale and

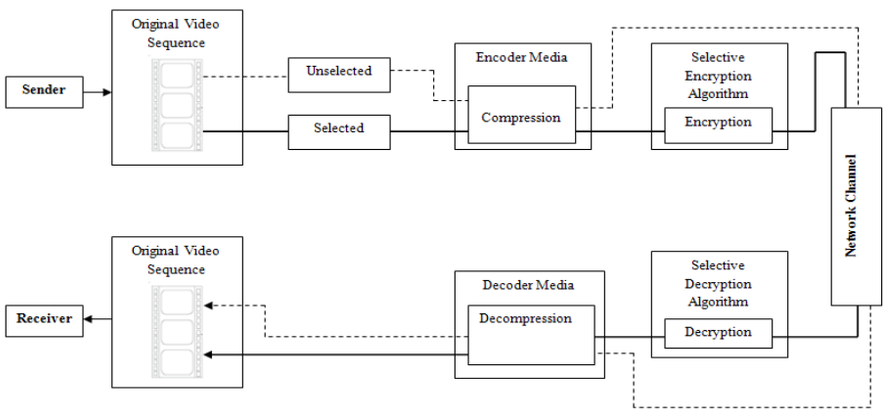


Figure 2. Structure of selective encryption algorithms.

strong points of correlation. However, for the video data, only a few pixels would differ between frames since the consecutive frames are similar[12]. Light encryptions are also important for use in many real-life applications in order for them to effectively preserve perceptual information[13]. The reasons are evident in different aspects of the proposed algorithm.

In this paper, a video encryption algorithm initiated from the audio samples that generates a public key that eliminates the need for distribution and management by a third party. The transferred digital multimedia is used to generate the public key over the internet. The proposed algorithm is efficient because it takes less time for encryption to be done. The encryption algorithms implement numerous iterations of transformation and substitution over the original data as in the plaintext. This helps in the complication of the process of data discovery by hackers or online intruders. The algorithm proposed in the paper can split the video into discrete segments; the frames and the samples of the audio. The initialization process is also done that involves sending a copy of the algorithm from the image to the receiver with streaming of bits for connection as per the established purpose. The proposed paper will be conducted to establish in details the subject of video encryption in comparison with audio and text encryptions.

2. Related Work

The previous section justified the need to examine multimedia technology. Because of the huge size of the digital videos, compressed formats are generally used for video transmission (i.e. MPEG-1, MPEG-2, MPEG-4, H.263, H.264, and AVC)[6-8]. Literature proposes different encryption approaches as possible solutions to protect digital images and videos. In this section, an extensive review of the literature is offered in particular with regards to different encryption methodologies and approaches that have been regularly used to make video communication more secure and reliable.

The first approach is the Naive Approach for video encryption proposed by Agi and Gong. In this approach, the video bit sequence stream is treated as text data and encrypted using symmetric key cryptosystem. However, even the fastest modern symmetric schemes such as DES or AES are computationally very expensive. Despite offering the highest level of security, they are not so well suited for video data encryption because they require large volume of data processing in real time[14].

The second approach was proposed by Gadegast, that is, the selective video encryption after the compression model, known as secure MPEG (SECMPEG). The after compression processes, SECMPEG, select a partial video stream or the whole video stream for encryption using light weight cryptographic algorithm. It then encrypts important part of the video using conventional encryption algorithm[9]. In 1995, Maples and Spanos proposed a similar approach to secure MPEG-I and MPEG-2. In this approach all I-frames but P-and B- in the MPEG video stream are encrypted. The MPEG video sequence header, that has important information for decoding process, is also encrypted. However, encrypting only I-frames has limited applications such as to the military where each and every part of the video data is important[15-17].

Based on the Shannon principle of diffusion and confusion, Choon[13] proposed a light weight and cost effective encryption approach where each picture is divided into macro-blocks. A macro-block is a 16 x 16 pixel array. By permutating macro-blocks followed by XOR operation on the permuted macro-block, the diffusion and confusion principles can be achieved. Another light-weight encryption approach on the uncompressed raw MPEG data was developed by Choo[14]. This approach is called Secure Real-time Media Transmission (SRMT). It uses two block transpositions and a XOR operation. However, Hooda and Parvinder made a survey on light-weight and cost effective encryption approaches[6].

Tang offered Zigzag permutation algorithm by embedding the encryption into the MPEG compression process. In this algorithm, a random permutation matrix that acts as a secret key is used to modify the ordering of transformation coefficient. In this scheme I-frames of MPEG video undergo zigzag reordering of 8 x 8 block to 1 x 64 vectors. There are three stages in this technique: The first stage generates a list of 64 permutations. The second stage involves splitting of 8 x 8 block. This is done by splitting the DC coefficient (8 bits) into two equal halves where 4 most significant bits are placed in DC coefficient and the least significant bits as the last AC coefficient. Finally, the then random permutation is applied to the split block[17].

A Video Encryption Algorithm (VEA) which works on compressed video stream was developed by Quio et al. The algorithm is based on statistical analysis. The basic idea of VEA is byte scrambling algorithm on output video data stream. It handles I-frames at

the slice level and process them bitwise. The data is divided into two byte stream as odd and even numbered bytes. These two streams are XORed to form the first part of the cipher. By performing DES over the even numbered byte streams, the second part of the cipher is constructed[4].

A puzzle encryption algorithm for compressed video stream is another approach to video encryption. Based on children's game puzzle, it was developed by Liu and Koeing[2]. This algorithm works in two steps: (1) the compressed video data of each frame is puzzled; and (2) the puzzled video data is then obscuring. This algorithm dramatically reduces the computational cost for video encryption as the encryption speed is sufficiently high to meet the real time requirements of mostly used multimedia applications, especially high resolution video games[18].

Wen et al. founded the selective encryption into format-compliant method, known as Compliant Configurable encryption format. In this scheme, data is first grouped into information carrying and non-information carrying parts. After that, only information carrying fields are encrypted. Field Length Code (FLC) or Variable Length Code (VLC) can be used to fix the information field. For format compliance, bits for encryption chosen and after encrypting with DES placed back to its original bit position in video stream[9-21].

Alattar et al. proposed three methods for selective video encryption based on DES cryptosystem. In the first method every n^{th} I – macro-block is encrypted, followed by the encryption of the headers of all the predicated macro-blocks and n^{th} macro-block data. In the third method, the n^{th} macro-block and the header of every n^{th} predicate macro-block are encrypted. This scheme works during compression[6].

In 2000, Cheng and Li offered partial encryption schemes for still images and they further extended them to the video. The partial encryption schemes works with quad-tree compression algorithms and wavelet compression algorithm based on zero-trees for the video stream I-frame, motion compensation, and residual error coding. This scheme works for the video stream based on Set Partitioning in Hierarchical Trees image compression algorithm. Because the proposed methods are not suitable for JPEG images, they are hence not applicable to MPEG video compression standard. Proposed partial encryption encrypts the I-frames, motion vectors and residual error code of video stream[6,7].

Daniel et al. proposed a novel video encryption algorithm specially designed for both lossless and lossy low motion spatial only video codecs. This algorithm works before the compression and at the receiver side after decompression. This is a unique feature and often desirable feature. However, it works only for certain classes of video sequence and codec[6,22,23]. This scheme works on the principle based on canonical sorting permutation σi of frame *Fi*. In this approach, canonical sorting permutation σ1 is computed for *F1* (first frame of video sequence *(F=F1, F2...,Fn)* and after compression *(C (F1))* is transmitted through secure channel without encryption. This first frame works as a secret key for the encryption and decryption process. Each subsequent frame *Fi* is encrypted by applying canonical sorting permutation $\sigma i\text{-}1(Fi)$. The receiver computes the sorting permutation for the received frame and uses it to recover the next frames from the encrypted frames.

Liu and Koenig stated that all video encryption algorithms have a relatively low security level although they meet the specific requirement of video encryption better than the naïve algorithm except for the security strength[6,18].

3. Methodology

~~Usually the data within a network~~ environment must be kept secure in concerns, storage and transmission since it is the most precious asset. Therefore, we should keep this data away from any unwarranted access by an unauthorized party. The concept that calls attention to the importance of creating and managing the keys is the cryptography[24]. In this paper, the security process performs encryption and decryption process for each video packet at both sender and receiver locations:

3.1 Encryption Process

Here, we describe the encryption process of the new proposed algorithm. Typically, the encryption processes are usually involved and engaged within the cryptographic process. The encryption algorithms include some preprocessing steps over the source data. For instance, the encryption algorithms execute many rounds of substitutions and transformations over the intended data[26]. These steps can efficiently decrease the probability of hacking and intruding.

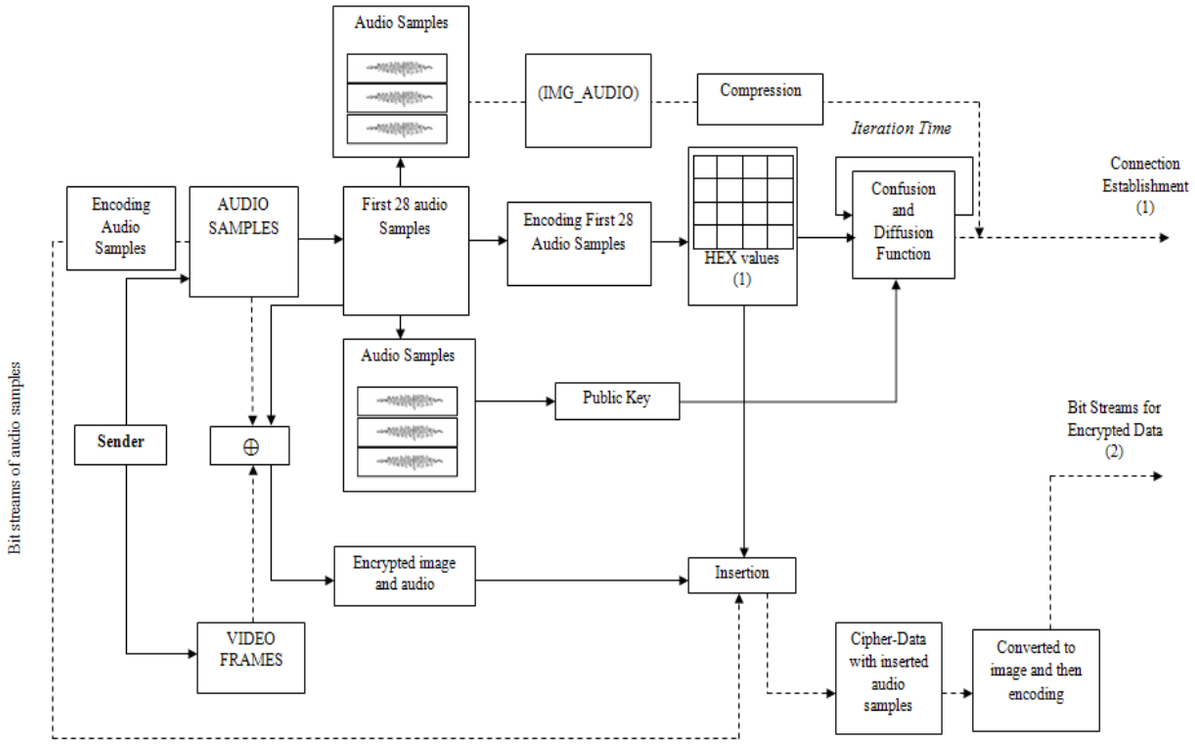


Figure 3. Flowchart of the proposed encryption algorithm.

As shown in Figure 3. The proposed encryption algorithm has many steps, starting from reading the defined video and ending with decrypting the video with synchronization process. Now, we will describe each step in more details:

Step 1: The user defined video reading process is done.

Step 2: Connection establishing step. In this step, the video is split into two dissimilar parts (audio samples, and video frames). The 28 audio samples are then gathered. Next, the bits stream is split also into 4 bits/units using the hexadecimal representation with 2D array i, j. The receiver of the data then take delivery of the first three audio samples which combined as image (i.e. Image of audio) after multiplying them with a window function. Typically, the frequency resolution has been decreased by using the window function.

Here, the Hamming window is distinguished by the following equation:

$$(W(n) = 0.54 - 0.46 \cos((2n\pi)/(N-2)),$$

Where, n= 0,…N. Then, the peak value is computed by the algorithm in order to produce the public key.

Step 3: Here, the complex two-dimensional array is generated based on the public key which generated from the previous step through confusion, diffusion, and repeated iteration. The confusion and diffusion is distinguished by the following equation:

$$Q = P\ D(C(Q, K_1), K_2))^n$$

Where, Q is the first 28 audio samples, and P represent the first 28 audio samples after confusion diffusion process, K_1 is confusion key, K_2 is diffusion key, and n is iteration time. *C()* and *D()* mean confusion process and diffusion process respectively.

Step 4: Here, in this step, the encrypted video frame is generated by using XOR procedure. XOR has been implemented over the set of audio samples with video frame in order to encrypt them. Moreover, the XOR process has the ability of achieving and completing the encryption process with minimal time, and it does not have an effect on the complexity of the encrypted data, however, it creates an extremely complex ciphered data because of its simplicity.

Step 5: In this step, the encrypted image has been modified by adding the audio sample inside of it, in order to make the confusion and diffusion ratio to be augmented

and increased. Typically, this step is mainly depends on step 2 output.

3.2 Decryption Process

Usually, each encryption process has a corresponding decryption process with a well defined keys and steps that are implemented over any important data (i.e. video stream). In the decryption process, all the data which altered by the encryption process to be un-comprehendible by any user, will be changed and modified to its original form as shown in Figure 4. Knowing the key (i.e. Public decryption key) which needed for completing the decryption process is considered as main principle for decrypting a ciphered data.

Thus, while the decryption process, the procedure of extracting the key correctly is considered as very crucial and important step. In the proposed algorithm, the key as mentioned before is generated from first 3 audio samples to the decryption process. Here, the decryption process has the following steps:

Step 1: Typically, the process of connection founding is done at the beginning of the encryption and decryption process, both of the sender and the receiver sides has this process. This process involves convert image of audio to audio samples (i.e. image to audio signals), next, it multiplied by a window function to produce public key, and create the HEX values that identical to the values of the sender side.

Step 2: Here, the HEX values are used to as a base to divide the encrypted data from the audio samples in order to complete the extraction process.

Step 3: In this step, the XOR is implemented over the extracted audio samples along with the encrypted data, after implementing this process; the original video frame will be ready.

Step 4: As a final step, the audio samples to video frames synchronization process is achieved to maximizing tolerances skew.

4. Experimental Results

The video data which was used for analysis have different motion characteristics and varying resolution with a frame rate of 30 fps. The sample test video sequences include videos like (i.e. Red 2010, and Fast and Furious). Some of the test videos along with their frame numbers are shown in Figure 5. Figure 6 shows the images of the corresponding video frames after applying XOR operation with audio samples. The corresponding images for some

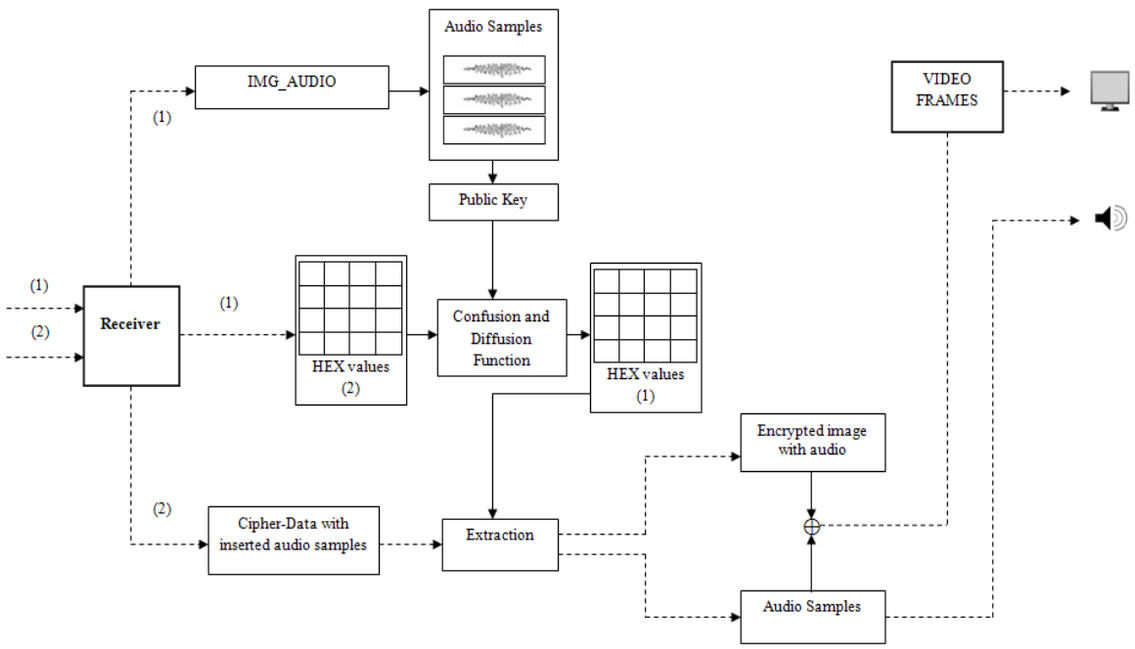


Figure 4. Flowchart of the proposed decryption algorithm.

of the test videos along with their frame numbers after applying the proposed algorithm are shown in Figure 7. It can be observed that the details in the video block are completely lost in the encryption process. The time spent for the pre-processing step in the proposed algorithm (i.e. connection establishment) is on an average 0.09003 ms per video. Since this is done infrequently, which is does not affect the encryption time of the video.

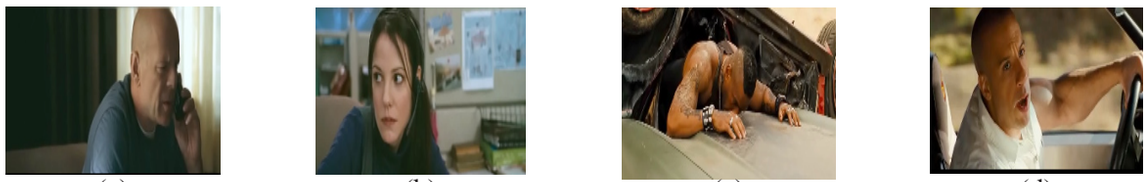
The experiments of the proposed encryption algorithm were implemented through MATLAB application tool version 7.0.0.1 using the DALI library, which supports coding of MPEG-1 and MPEG-2 videos[27] on a 1.6 GHz core i5 (IV), 8 GB memory and 750 GB hard disk capacities.

5. Security Analysis

This paper aims to propose a new applicable and secure video encryption algorithm to improve video encryption performance by minimizing the security level. This section presents the security analysis. In order to test the security of the proposed algorithm, a set of tests and analysis was performed on the proposed algorithm. The security analysis of the proposed algorithm was conducted in two phases: Audio samples distribution phase and reconstruction phase.

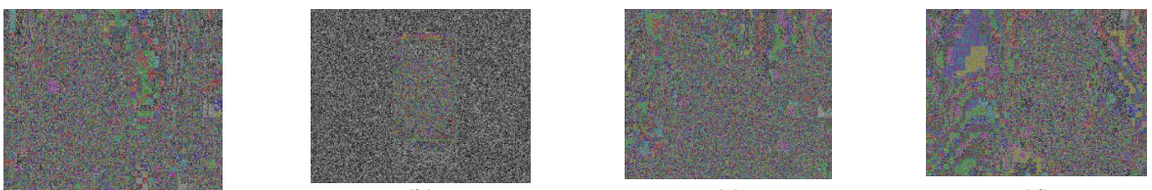
5.1 Audio Samples Distribution Phase

As mentioned earlier in the proposed algorithm (i.e. Encryption Process), the complex two dimensional array is generated through a regular confusion, diffusion, and repeated iteration based on the public key which generated from the audio samples. Here, we have two arrays; the first one is holding the data before performing a regular confusion, diffusion, and repeated iteration for insertion audio samples, where the second one is yielded after performing a regular confusion, diffusion, and repeated iteration. To examine the relation between these two arrays, correlation distribution analysis was performed as shown in Figure 8.



(a) (b) (c) (d)

Figure 5. Some of the test videos along with their frame numbers before applying the proposed algorithm.

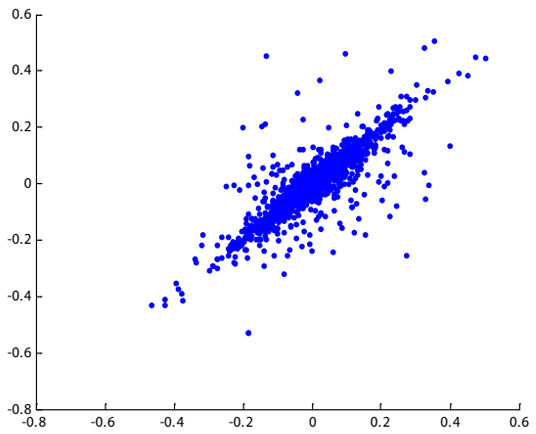


(a) (b) (c) (d)

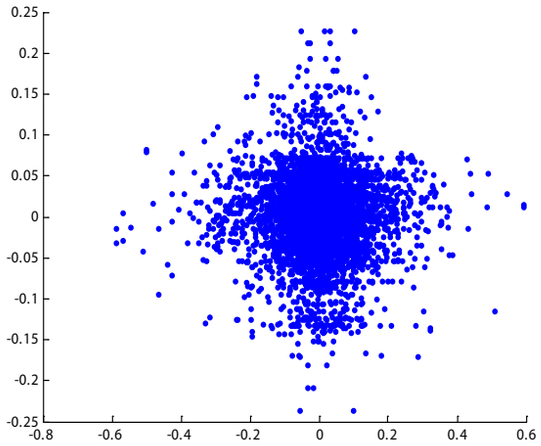
Figure 6. Corresponding images for some of the test videos along with their frame numbers after applying XOR operation.

(a) (b) (c) (d)

Figure 7. Corresponding images for some of the test videos along with their frame numbers after applying the proposed encryption algorithm.



(a) Correlation distribution before applying confusion and diffusion function.



(b) Correlation distribution after applying confusion and diffusion function.

Figure 8. Correlation distribution analysis.

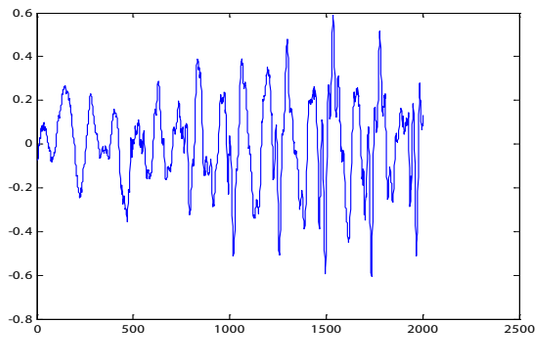


Figure 9. Audio samples around zero.

It is clear from Figure 8b that confusion and diffusion shown good performance but most of the values are close to zero due to the high audio frequency (32 KHz) as shown in Figure 9. Thus, the correlation distribution after confusion and fusion process will be randomly distributed around zero.

5.2 Reconstruction Phase

In general, the encrypted frames are then decrypted by applying the inverse transformation function (i.e. decryption process). Thus, the original frames are reconstructed. In this paper, the performance analysis for the reconstruction process is measured by means of parameters such as Peak Signal to Noise Ratio (PSNR) and Mean Squared Error (MSE)[28]. MSE and PSNR are distinguished by the equations 1, and 2 respectively:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j)-k(i,j)\right]^2 \quad (1)$$

And,

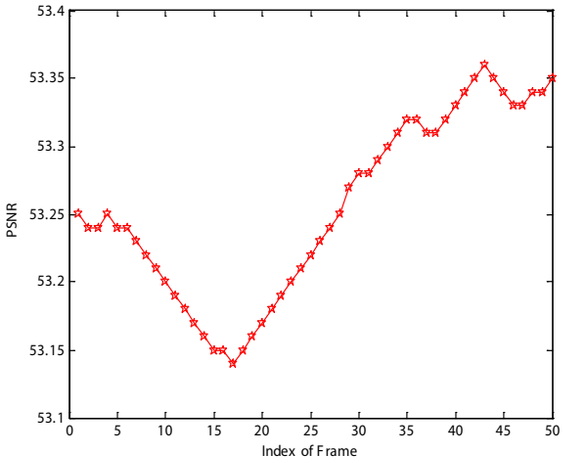
$$PSNR = 20.\log_{10}\left(MAX1\right) - 10.\log_{10}\left(MSE\right) \quad (2)$$

Where, MAX1 is the maximum possible pixel value of the image. Typical values for the PSNR in the decrypted image and video compression are between 30 and 58 dB, where higher is better, and the typical values for the MSE is equal or close to zero[28]. The corresponding distribution of the proposed algorithm for both PSNR and MSE are shown in Figure 10.

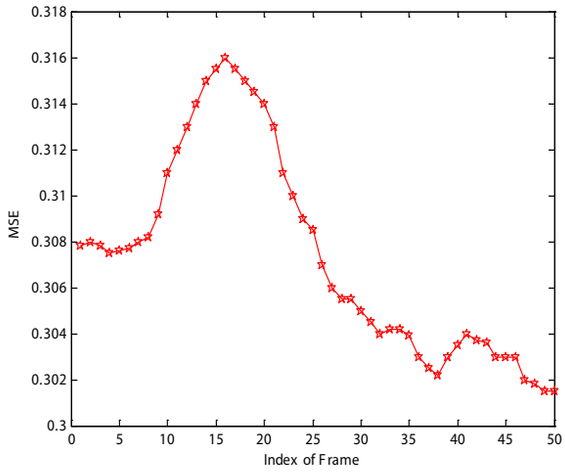
It is clear from Figure 10 that the PSNR values for the proposed algorithm (i.e. 50 frames) within the typical values, while the MES values are close to zero. This indicates that the proposed algorithm shown good performance.

6. Conclusion

In this paper, we have achieved our objectives of designing a new cryptographic encryption algorithm by combining audio samples and image frames for enhancing video security. Most of the available encryption algorithms are not suitable to be used to over an open network since



(a) PSNR analysis of the video according to the proposed algorithm.



(b) MES analysis of the video according to the proposed algorithm.

Figure 10. Reconstruction analysis.

they were originally built for binary and text data and due to their extensive computations which result a significant overhead and may be too expensive for real-time applications such as image surveillance and video conferencing. Thus, the two criteria's such as transmission speed and higher security level are important for the successful transmission of the video data. In this paper, a Simple XOR operation is used to encrypt the video data in order to minimize the computation cost, and combining audio samples and image frames in a random fashion in order to enhance video security. The conducted experiments and the security analysis of the proposed algorithm show that the algorithm is strong, fast, and secure because it satisfied the standard security analysis such as correlation analysis. Moreover, the performance of the proposed algorithm is measured by means of parameters such as Peak Signal to Noise Ratio (PSNR) and Mean Squared Error (MSE).

7. References

1. Thomas E, Kumar S, Paar C, Poschmann A, Uhsadel L. A survey of lightweight-cryptography implementations, IEEE Design and Test of Computers. 2007; 24(6):522–33.
2. Liu F, Koenig H. A survey of video encryption algorithms. Computers and Security. 2010; 29(1):3–15.
3. Rajanbabu DT, Raj C. Multi level encryption and decryption tool for secure administrator login over the network. Indian Journal of Science and Technology. 2014; 7(4):8–14.
4. Qiao L, Klara N. A new algorithm for MPEG video encryption. Proceedings of First International Conference on Imaging Science System and Technology. 1997; p. 21–9.
5. Ramalingam M, Isa NAM. A steganography approach over video images to improve security. Indian Journal of Science and Technology. Jan 2015; 8(1):79–86.
6. Hooda D, Singh P. A comprehensive survey of video encryption algorithms. IJCA. 2012; 59(1):14–9.
7. Valarmathi R. Secure data transfer through audio signal with LSA. Indian Journal of Science and Technology. 2015; 8(1):17–22.
8. Eskicioglu AM, Delp EJ. An overview of multimedia content protection in consumer electronics devices. Signal Processing: Image Communication. 2001; 16(7):681–99.
9. Meyer J, Gadegast F. Security mechanisms for multimedia data with the example MPEG-1 video. Project Description of SECMPEG, Technical University of Berlin, German; 1995.
10. Liu X, Eskicioglu AM. Selective encryption of multimedia content in distribution networks: Challenges and new directions. IASTED Communications, Internet and Information Technology (CIIT), USA; 2003.
11. Kunkelmann T. Applying encryption to video communication. Proceedings of the Multimedia and Security Workshop at ACM Multimedia. 1998; 98:41–7.
12. Socek D, Hari K, Magliveras SS, Marques O, Culibrk D, Furht B. New approaches to encryption and steganography for digital videos. Multimedia Systems. 2007; 13(3):191–204.
13. Choon LS, Samsudin A, Budiarto R. Lightweight and cost-effective MPEG video encryption. Proceedings of

Information and Communication Technologies: From Theory to Applications. 2004; p. 525–26.
14. Choo E, Lee J, Lee H, Nam G. SRMT: A lightweight encryption scheme for secure real-time multimedia transmission. IEEE International Conference on Multimedia and Ubiquitous Engineering, MUE'07; 2007. p. 60–5.
15. Massoudi A, Lefebvre F, De Vleeschouwer C, Macq B, Quisquater JJ. Overview on selective encryption of image and video: challenges and perspectives. EURASIP Journal on Information Security; 2008.
16. Qiao Land Klara N. Comparison of MPEG encryption algorithms. Computers and Graphics. 1998; 22(4):437–48.
17. Tang L.Methods for encrypting and decrypting MPEG video data efficiently. Proceedings of the Fourth ACM International Conference on Multimedia, ACM. 1997; p. 219–29.
18. Liu F, Koenig H. Puzzle-an efficient, compression independent video encryption algorithm. Multimedia Tools and Applications. 2014; 73(2):715–35.
19. Zeng W, Lei S. Efficient frequency domain selective scrambling of digital video. IEEE Transactions on Multimedia. 2003; 5(1):118–29.
20. Bhargava B, Shi C, Wang SY. MPEG video encryption algorithms. Multimedia Tools and Applications. 2004; 24(1): 57–79.
21. Wu CP, Kuo CCJ. Design of integrated multimedia compression and encryption systems. IEEE Transactions on Multimedia. 2005; 7(5):828–39.
22. Socek D, Kalva H, Magliveras SS, Marques O, Culibrk D, Furht B. New approaches to encryption and steganography for digital videos. Multimedia Systems. 2007; 13(3):191–204.
23. Wang Z, Bovik AC. A universal image quality index. IEEE Signal Processing Letters. 2002; 9(3):81–4.
24. 24. Al-hazaimeh OMA. Increase the security level for real-time application using new key management solution. International Journal of Computer Science Issues (IJCSI). 2012; 9(3):240–46.
25. Sasi SB, Sivanandam N. A survey on cryptography using optimization algorithms in WSNs. Indian Journal of Science and Technology. 2015; 8(3):216–21.
26. Al-hazaimeh OMA. A novel encryption scheme for digital image-based on one dimensional logistic map. Computer and Information Science. 2014; and 7(4):65.
27. Acharya B, Nikhil T, Arasu DR, Vishnu Prasad N. Encryption and decryption of informative image by key image using modified Hill cipher technique based on non-invertible matrices. Proceedings of the 2011 International Conference on Communication, Computing and Security. p. 606–09. ACM; 2011.
28. Tao S, Wang R, Yan Y. Clock-controlled chaotic keystream generators. Electronics Letters. 1998; 34(20):1932–4.