ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20240924002013
Original scientific paper

Enhanced Video Encryption Technique Using Substitution Boxes and Spatial Filtering for Robust Data Security

R. SAKTHEESWARI, C. YAASHUWANTH*, K. PRATHIBANANDHI, S. RAMESH

Abstract: Video encryption is a technique used to protect video data, which has become increasingly crucial due to the rise of intruders attempting to steal private information. Ensuring the protection of sensitive data during transmission is of paramount importance across various domains. Current encryption algorithms often struggle to balance security and efficiency, resulting in compromises on one or both aspects. To address this issue, we propose an enhanced video encryption technique utilizing substitution boxes (S-boxes) and spatial filtering. The proposed method avoids the overhead caused by the repetition of chaotic sequences, leading to improved encryption efficiency. Authentic video frames, which are unaltered, are converted into three-dimensional videos. Spatial permutations, spatial filtering, and image superposition are applied in three rounds along the x, y, and z axes. S-boxes are generated to provide an additional layer of security. Throughout the process, pixels in the video frames are randomly switched to ensure the final encrypted video is secure. Compared to currently available algorithms, the results of our analysis demonstrate that the proposed method achieves superior levels of both security and efficiency, making it well-suited for ensuring robust data security during transmission.

Keywords: data security; chaotic sequence optimization; spatial filtering; substitution box (s-box); video encryption;

1 INTRODUCTION

The field of cryptography is the one to turn to when it comes to protecting sensitive data from being stolen by unauthorized individuals. To achieve this objective, it is imperative to convert the data into an encrypted format that is unintelligible and inaccessible to unauthorized individuals during transmission or storage. [1]. Encrypting data is the primary objective of data cryptography, which aims to render information unintelligible or invisible while it is being transmitted or stored. It encompasses not only text but also images, audio, and video. The objective of cryptography is to safeguard confidential data and maintain its integrity by thwarting potential hackers and other malicious entities from unauthorized access. Decryption of data refers to the process of reversing the encryption process in order to retrieve the data in its original form. Cryptography has undergone significant changes since its early days of use in ancient Egypt. These changes have been influenced by significant events that have altered the way people handle data [2]. By deciphering the Enigma cipher machine, which was used by the Germans to encrypt military communications, the Allies were able to gain a decisive advantage and significantly contribute to the speedy conclusion of the war [3].

In many contemporary real-time applications, video encryption is utilized to a significant degree. The use of drones, also known as unmanned aerial vehicles, is beneficial in a variety of situations where the transmission or recording of video is required. Video data that has been captured and stored by drones is sent to a control center for processing. The control center is able to perform additional analysis on the data, which can then be utilized for the purposes of public safety and Intelligent Transportation Systems (ITS). These films are available to a wide range of audiences, including members of the law enforcement and emergency services communities, among others [4]. Due to the fact that the video may contain sensitive information, including the faces of onlookers, which may raise concerns about privacy, encryption becomes necessary in this context. In the modern era of the internet, there has been a significant increase in the amount of digital content, particularly video, that is available on a variety of online platforms. As a result, this technology also has applications in the copyrights sector. There are additional applications for drones in the military, healthcare, satellite, and space industries, in addition to the applications that are already in place [5]. A video is made up of moving still images, which are referred to as frames, that are displayed one after the other in rapid succession. Because security measures are not as stringent as they should be, users of the internet are increasingly susceptible to having their personal information (which may include text, photos, audio, and video, among other things) intercepted and stolen. It is of the utmost importance to ensure the safety of multimedia files because of the significant amount of sensitive information that they contain. Multiple security measures have been implemented by the government in order to ensure that data transmissions over the internet are protected from unauthorized access [6]. These include digital signatures, authentication, and encryption mechanisms.

The use of encryption to safeguard confidential information is a worthwhile endeavor. At the moment, a significant amount of research attention is being directed toward the creation of methods that are both effective and safe for encrypting video and image files [7-10]. The encryption of video was not the primary objective of traditional encryption algorithms such as AES, RSA, IDEA, and DES when they were first developed. Unfortunately, due to the fact that they are so slow and require a lot of resources, these algorithms are not suitable for encrypting videos in real time.

2 LITERATURE SURVEY

For electronic financial transactions, resource restriction, trust establishment, and information security, cryptography is the gold standard and an essential part of any company's security policy. Its original purpose of protecting sensitive military information has expanded. Data that is not encrypted and is meant to be transmitted or stored is called plaintext. Anyone, computer or not, should have no trouble reading and understanding it. Both humans and computers are unable to interpret or process ciphertext until it is decrypted. Any device or system that can encrypt and decrypt data is called a cryptosystem [11]. The complexity of an encryption process, the software required

to implement it, and the key - typically a long sequence of bits - used to encrypt and decrypt data are all defined by the cryptosystem's encryption algorithm. A famous theory about the first security principle of all encryption systems was put forth by Kerchhoff in the nineteenth century. Experts in the field of cryptosystem development now consider this theory to be fundamental. Kirchhoff made the astute observation that adversaries are likely to be aware of the encryption algorithms [12]. The encryption system will stop protecting the ciphertext once the algorithm is deciphered, regardless of whether the adversary knows this at first or not. Examining the size of the key space produced by an encryption algorithm is one method for determining how secure the algorithm might be [13]. The larger the key space, the more time it will take an attacker to search it thoroughly; hence, the higher the level of security. When discussing encryption, a key is defined as the unique piece of data that dictates how plaintext is transformed into ciphertext or reversed during decryption. Key values are often long strings of random bits. The keyspace is a range of values that must be combined in order to generate an encryption key. More possible keys can be generated with a larger keyspace. Key sizes of 128 bits, 192 bits, or 256 bits are typical in contemporary cryptography, for instance. Thus, the keyspace would be 2^256 with a 256-bit key [14].

The key's secrecy, key length, initialization vector, and their interplay define an encryption algorithm's resilience. Key length and encryption algorithm both play important roles in determining the encryption strength. A cipher is deemed weak if it can be cracked in three hours using only a Pentium 4 processor [15]. The cipher is strong if cracking it using thousands of multiprocessing systems takes a million years. An extensive evaluation and synthesis of all pertinent literature on the selected research subject is an essential component of any literature review. In order to guarantee the safety of video data, this review summarizes articles that offer solutions.

Using key frames and a standard sorting permutation list, this paper demonstrates a method for digital video encryption [16]. When encrypting data sent between frames, this paper mostly uses the xor operation, which stands for bitwise exclusive OR. In this particular setting, two essential visuals are used as covert symbols, and the frames are arranged in a methodical manner by employing a conventional permutation structure. The metrics that are utilized in the process of determining the quality of each video are the Peak Signal to Noise Ratio (PSNR), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Not only does this paper include the video frames that take place before and after encryption, but it also provides a histogram for both of those frames . This method improves the efficiency and safety of the system's cryptography, which is a major plus. To encrypt video frames, the author first proposed the Huffman algorithm. Before the video is encrypted using the block cipher technique, it is first divided into multiple frames. The frames are then converted into blocks. The time required by selective methods is far lower than that of full encryption. Encoding cannot take place without secret keys [17-19].

An approach to deciphering encryption that works with videos and images is presented by the author. While improving scalability and security, selective encryption algorithms reduce the data quality that needs to be encrypted. The effectiveness of the selective encryption algorithm is evaluated using multiple criteria. Before, during, and after compression are the three main types of selective encryption. These groups are defined by the timing of compression during encryption and have few real-world applications. On every occasion, when it comes to selective encryption, the compression method is way better.

3 PROPOSED WORK

The video is sourced from the telemedicine sessions, patient consultations and surgical procedure. The video might contain crucial information that has to be secured. The private key is converted into the two-dimensional logistic modular map. Following this the chaotic sequence is formed and the SHA 256 hash value is applied for providing security. The input video is transformed into three-dimensional space. The plane level permutation in the x axis is carried out. Plane level permutation refers to the ordering of the pixels in a specific arrangement. This is done within the plane while the structure of the image is kept unaltered. The next step involves the plane level video frame filtration in the z axis. Then the image super position is done. Fig. 1 shows Block diagram of the proposed method.

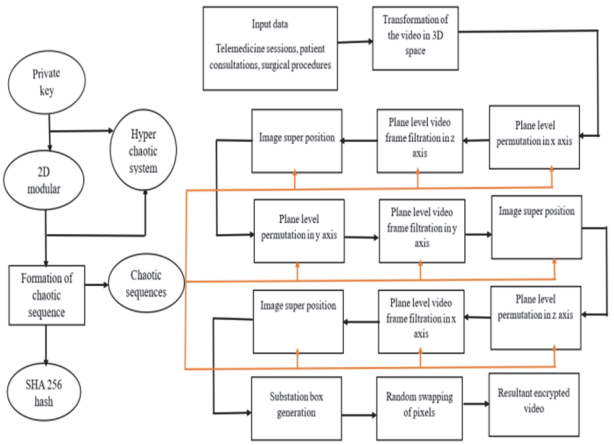

Figure 1 Block diagram of the proposed method

The above procedure is now repeated for the *y* and *z* axis. The plane level permutation in the *y* axis is calculated following which the plane level video filtration in the y axis is carried out. After performing the same operations of video frame filtration ns plane level permutation in the *z* axis, the substitution box is generated to offer security during the video encryption. The pixels are swapped in a random manner to further improve the security of the encrypted video. Eventually the resultant image is obtained.

3.1 Chaotic System

The chaos in the high dimensional system is much more complex than that of the low dimensional system. Thus, it is mostly preferred by the designers and the researchers. The chaotic system in the four-dimensional space is represented by the following equation.

$$\begin{cases} a = \alpha(b-c) \\ b = \lambda a - ac + ub - w \\ c = ab - \beta c \\ w = a + \gamma \end{cases} \quad (1)$$

Here a, b, c, w are the non-constants of the state. The parameters that control the system are α, β, u, λ. When α = 36, β = 3, u =28, λ = −16, the system is said to be hyper chaotic. The enhanced features of the 4-dimensional chaotic system makes it more preferable for the video encryption techniques.

$$a_{i+1} = \begin{cases} \left(4s_1 a_1 (1-a_i) + 2s_2 b_i\right) \bmod 1; & \text{when } b_i < 0.5 \\ \left(4s_1 a_1 (1-a_i) + 2s_2 (1-b_i)\right) \bmod 1; & \text{when } b_i \geq 0.5 \end{cases} \quad (2)$$

$$b_{i+1} = \begin{cases} \left(4s_1 b_1 (1-b_i) + 2s_2 a_i\right) \bmod 1; & \text{when } b_i < 0.5 \\ \left(4s_1 b_1 (1-b_i) + 2s_2 (1-a_i)\right) \bmod 1; & \text{when } b_i \geq 0.5 \end{cases} \quad (3)$$

Here the outputs are represented by a_i and b_i for the ith iteration and it also represents the inputs of the next iteration of i. the parameters of the logistic modular map of 2 dimension are s_1 and s_2. When s_1 and s_2 lies between 1 and 100, the 2-dimensional logistic modular map is called as the hyperchaotic state.

3.2 SHA 256 Hash Value

This technology was initially introduced by the NIST in the year 2001. This technology has now become more prevalent in the block chain technology. This is because of its reasonably good performance.

$$\begin{cases} H_1 = \left(\sum_{i=1}^{32} H_i\right) \bmod M \\ H_2 = \left(\sum_{i=1}^{32} H_i\right) \bmod N \end{cases} \quad (4)$$

It is a hash function that is based on cryptography. This sha 256 algorithm is not directly used for the encryption of the video, but it is used for the generation of a hash value that has a fixed size of 250 bits from the data that is fed as an input.

It is efficient because of its integrity and its authentication. The hash value is generated by passing the video into the sha 256 hash function. This hash value turns out to be the unique identification of that video. This hash value can be obtained during the decryption of the video and this in turn will help the user to check the integrity of the video.

This hash function can also be used to check if the video has been sent by the intended user. When the sender sends the video, the sender can encrypt the video and the hash function using a private key. The decryption can be done by the receiver by using the key given by sender. This helps the users to identify if the data has been sent by the intended user or not. The cryptographic passwords can be generated by using the sha 256 function. The encryption using symmetric algorithms can be done by using the produced keys. This also finds its application in the creation of digital signatures.

3.3 S Box Generator

The existing S box generators are not suitable for the small level cryptography techniques. Hence in the proposed work, an ordered elliptic curve is used for the generation of the random integers in the range of 0 to 2n −1. The ordered elliptical curve is also used for the generation of the binary sequence. The binary sequences are used to create the substitution boxes that are dependent on the plain text.

The proposed substitution box generation has the following steps involves in it.

Step 1: Two sequences are chosen

$X^i = \left(x_0^i, x_1^i, ..., x_l^i\right)$ within the set of positive integers (i.e) i = 1, 2, 3…..

Step 2: An elliptical curve is chosen such that $q \equiv 2|3|$, $q \geq 2^n$

Step 3: Teo different sets are chosen such that $Y^i = \left\{y_0^i, y_1^i, ..., y_l^i\right\}, \left|Y^i\right| = 2^n, Y^i|2| = [0,l]$

Step 4: Perform computation of the sets $Z^i = \left\{z_0^i, z_1^i, ..., z_l^i\right\}, z_j^i = \left(a_j^i, y_j^i + x_j^i |p| \in E_q\right)$

Step 5: The unevenness in Y^i is generated with respect to Z^i in a way that y_j^i is smaller.

Step 6: from the ordered Y^i, let us assume K_i to be the sequence obtained. It is denoted by $K^i = \left(k_0^i, k_1^i, k_2^i ..., k_l^i\right)$ for $i = 1,2$

Step 7: Now the substitution box is generated $\ell^{1,2} = \left(X^i, T^i, Y^i, q, y, n\right) : [0,l]$

$$\ell^{1,2} = \left(X^i, T^i, Y^i, q, y, n\right)\left(k_j^1\right) = k_j^2, j \in [0,l]$$

Step 8: Create the binary sequence which has assize of n_2n and perform division starting from left side to right side. These sequences are then converted to decimal numbers.

Step 9: The order of the integers is defined within [0, l] depending on the value of N.

Step 10: the s boxes are generated for i = 1, 2, …. Such that $\ell^i = \left(X^i, T^i, Y^i, q, y, n\right)\left(p_j\right) = k_j^i, j \in [0,l]$

Determination of three dimensions

To start with, the plain image of the video must be converted from 2-dimensional space to 3-dimensional space. For the two-dimensional image, the following steps are carried out to convert it to the intermediate cipher image in the 3-dimensional space.

Step 1: The total count of the pixels is determined tP. If the value of this total count is not in the powers of 2 or if it is less than eight, P is stuffed with zeroes until the value of the total pixel count becomes equal to 2 or less than eight. If not, continue the procedure and go to the next step.

Step 2: Calculate the value of theta with the help of the calculated total pixel value.

$$\theta = \left\lfloor \log_2 (t)^{(S)} / 3 \right\rfloor$$

Step 3: The size of the initial dimension is found by using $D_{(1)} = 2^{\phi}$

Step 4: The second dimension can be determined through $D_{(2)} = D_{(1)}$

Step 5: According to the total pixel count, first and the second dimension, the size of the third dimension can be calculated.

$$D_3 = D^{\log_2(t)^{(S)} - D_1 - D_2}$$

Step 4: The P value is reshaped with $D_{(1)} \times D_{(2)} \times D_{(3)}$

A test image that having the size of 256 × 256 is converted as 32 × 32 × 64 in the three -dimensional space. In the similar way, other test images of sizes 512 × 512 and 1024 × 1024 will be converted as 64 × 64 × 64 and 64 × 64 × 256 in the three-dimensional space.

3.4 Generation of Chaotic Sequence

In the proposed work, the encryption of the video's plain image leverages the sequence of unevenness. The private key is converted to the beginning state values. Then along with the control parameters, these are fed as inputs to the system, for the generation of the chaotic sequence.

$$L_1 = H_1 + \rho_D$$
$$L_2 = \rho_D + 4\pi_D + 3$$

where, $\pi_D = D_1 \times D_2 \times D_3$

This in turn is converted into ten different sequences of chaos that can be used for the process of encryption.

$$\begin{cases} S_{(1)} = S_{(1)}\left(H_{(1)} + 1 : H_{(1)} + D_{(1)}\right) \\ S_{(4)} = S_{(1)}\left(H_{(1)} + D_{(1)} + 1 : H_{(1)} + D_{(1)} + D_{(2)}\right) \\ S_{(7)} = S_{(1)}\left(H_{(1)} + D_{(1)} + D_{(2)} + 1 : H_{(1)} + \rho_D\right) \end{cases} \quad (5)$$

$$\begin{cases} S_{(2)} = \left[\left|S_{(2)}\left(1 : D_{(3)} + 1\right)\right| \times 10^{17}\right] \bmod 256 \\ S_{(5)} = \left[\left|S_{(2)}\left(D_{(3)} + \pi_{(D)} + 2 : D_{(2)} + D_{(3)} + \pi_{(D)} + 2\right)\right| \times 10^{17}\right] \bmod 256 \\ S_{(8)} = \left[\left|S_{(2)}\left(D_{(2)} + D_{(3)} + 2\pi_{(D)} + 3 : \rho_{(D)} + 2\pi_{(D)} + 3\right)\right| \times 10^{17}\right] \bmod 256 \end{cases} \quad (6)$$

$$\begin{cases} S_{(3)} = \left[\left|S_{(2)} D_{(3)} + 2 : D_{(3)} + \pi_{(D)} + 1\right| \times 10^{15}\right] \bmod 256 \\ S_{(6)} = \left[\left|S_{(2)} D_{(2)} + D_{(3)} + \pi_{(D)} + 3 : D_{(3)} + 2\pi_{(D)} + 2\right| \times 10^{15}\right] \bmod 256 \\ S_{(9)} = \left[\left|S_{(2)} \rho_{(D)} + 3\pi_{(D)} + 4 : \rho_{(D)} + 4\pi_{(D)} + 3\right| \times 10^{15}\right] \bmod 256 \\ S_{(10)} = \left[\left|S_{(2)}(\rho_{(D)} + 3\pi_{(D)} + 4 : \rho_{(D)} + 4\pi_{(D)} + 3\right| \times 10^{15}\right] \bmod \pi_{(D)}) + 1 \end{cases} \quad (7)$$

$S_{(1)}, S_{(4)}, S_{(7)}$ can be used for performing the permutation operations related to the spatial domain, $S_{(2)}, S_{(5)}, S_{(8)}$ can be used for performing the filtering operations related to the spatial domain and $S_{(3)}, S_{(6)}, S_{(9)}$ are used for the generation of the matrices involving the chaotic pixels. $S_{(10)}$ is used for the swapping of the pixels in a random manner.

3.5 Permutation Operations in the Spatial Domain

Permutation refers to the operation which involves the rearrangement of the pixels in the same domain. This can be used for several applications such as data encryption, hiding and video image scrambling. Fig. 2 represents the permutation operation in the spatial domain.

The rearrangement of the pixel means, the position of the pixels change but the value of the pixels does not change. The algorithm related to this permutation defines how these pixels are rearranged. It needs a value of the key in several cases. The key is used to ensure that the permutation process can be revered if there arises a need and it provides privacy during the encryption of hiding of the data.

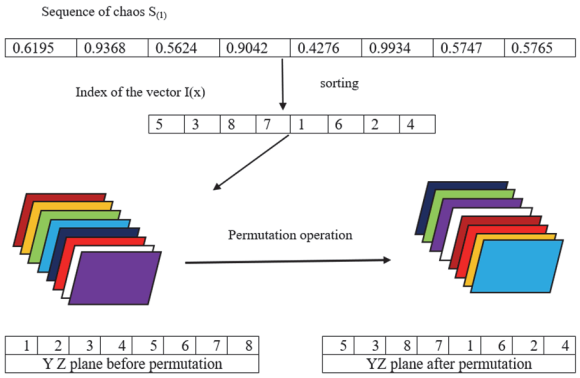

Figure 2 Permutation operation in the spatial domain

3.6 Video Filtering Operation in the Spatial Domain

The filtering is a common technique that can be used for removing noise in the image, detection of the edges and so on. Traditional filtering techniques cannot be applied directly on the encryption process since they cannot be reversed. Multiple pixels can be diffused into the current pixel through convolution technique. Fig. 3 gives the pixel level filtering of the video image.

$$C_9 = \left(\left(\sum_{j=1}^{N} m_i \times p_i + 1 \right) \times p_9 \right) \bmod 256 \quad (8)$$

Pixel level filtering process requires a mask of size 3 × 3 to diffuse in to the neighbouring pixels. The coefficients of convolution are represented by *k*.

$$P_9 = \left(\left(C_9 - \sum_{j=1}^{N} m_i \times p_i + 1 \right) \right) \bmod 256$$

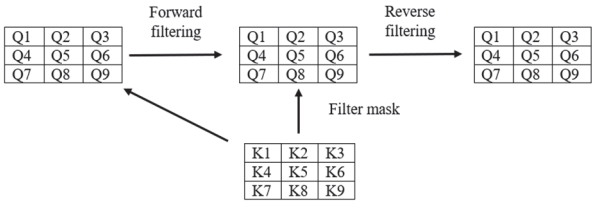

Figure 3 Pixel level filtering of the video image

Pixel-level image filtering diffusion is an effective technique, despite the fact that it diffuses pixels on an individual basis like other methods. The efficiency of this diffusion process can be further increased by using the proposed method.

$$C(*,*,5) = \left(\left(\sum_{j=1}^{4} m_i \times p(*,*,5) \right) + 1 \times p(*,*,5) \right) \bmod 256 \quad (9)$$

$$P(*,*,5) = \left(C(*,*,5) - \sum_{j=1}^{4} p(*,*,5) \right) \bmod 256 \quad (10)$$

The plane level filtering requires a mask of 1 × 5 to diffuse into the neighbouring pixels. The convolution coefficients are represented by *k*. The same mask can be used in the reverse filtering operation of the video images.

3. Performance Measurement of the Video Encryption Technique

Statistical measure:

The unevenness in the encrypted video frames can be calculated by using statistical measures. The vector correlation between *A* and *B* of the neighbouring pixels in every frame in the video can be determined by

$$conv(A,B) = \frac{1}{M} \sum_{i=1}^{M} \left(a_i - \frac{1}{M} \sum_{i=1}^{M} a_j \right) \left(b_i - \frac{1}{M} \sum_{i=1}^{M} b_j \right) \quad (11)$$

$$D(A) = \frac{1}{M-1} \sum_{i=1}^{M} \left(a_i - \frac{1}{M} \sum_{i=1}^{M} a_j \right)^2 \quad (12)$$

In order for a method of video encryption to function properly, there must be very little to no correlation between neighbouring pixels in both the vertical and horizontal directions.

Along with this, the histogram of the video encryption must be even and a little different from the original one. The sensitivity of the encryption technique can be measure by this difference between the original clip and the resultant clip.

Mean absolute error:

An efficient video encryption technique should produce video of that area little different from the actual video. The average variation in the intensities of the source video and the resultant video can be determined as follows.

$$Mean\ absolute\ error = \frac{1}{M \times M} \sum_{i=1}^{M} \sum_{j=1}^{M} \left| Q(i,j) - E(i,j) \right| \quad (13)$$

The original video frame's color intensity values are denoted by $Q(i, j)$ and the encrypted video frame's values are denoted by $E(i, j)$.

Mean square error:

The mean square error between the frames of the videos is given by

$$Mean\ sqauare\ error = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} \left| A(i,j) - B(i,j) \right|^2}{M \times M} \quad (14)$$

The similarity value will be large when the mean square value reaches zero.

Change of pixel rate:

The change in the rate of the pixel is used for the determination of the percentage of pixel variation between the actual and the resultant video.

$$D(i,j) = \begin{cases} 0, E_1(i,j) = E_2(i,j) \\ 1, E_1(i,j) \neq E_2(i,j) \end{cases} \quad (15)$$

$$Changing\ pixel\ count = \frac{1}{M \times M} \sum_{i=1}^{M} \sum_{j=1}^{M} D(i,j) \times 100\% \quad (16)$$

Here E_1 and E_2 are the frames of the video. The sensitivity of the video encryption technique can be measured by using this parameter.

Average intensity variation:

The variation in the average intensity between the two frames can be determined by using the following equation.

Average intensity variation

$$= \frac{1}{M \times M} \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{\left| E_1(i,j) - E_2(i,j) \right|}{255} \times 100\% \quad (17)$$

Similar to the change of the pixel rate calculation, here too E1 and E2 are the frames of the video. The sensitivity of the video encryption technique can be measured by using this parameter.

Quality measure during reconstruction:

For Evaluating the quality of the reconstruction by using the ratio of the peak signal to the noise. A logarithmic method is utilized in order to graphically represent the mean square error.

$$peak\ signal\ to\ noise\ ratio = 10 \log_{10} \frac{\max^2}{mean\ sqaure\ error} \quad (18)$$

Here max represents the maximum colour intensity value which is usually 255.

4 RESULTS AND ANALYSIS

The process of analyzing all of the possible permutations of the keys used in a cryptographic method is referred to as key space analysis. What is referred to as the key space is the number of different keys that an encryption algorithm is able to use. The size and the complexity involved in the encryption process if the video can be analyzed using the key space analysis. The occurrence of the brute force attack becomes difficult when the key space is large. The size of the key space also governs how strong the encryption process will be. The scalability of the S-box to larger video datasets or real-time applications depends on its computational efficiency and memory requirements. Precomputed lookup tables and parallel processing ensure fast substitutions, enabling real-time performance. Optimized S-box designs can handle high-resolution and large-scale video datasets effectively without significant latency.

$$\begin{cases} a_0^{(1)} = x_1 x_2 \ldots\ldots x_{52} \times 2^{-52} \\ b_0^{(1)} = x_{53} x_{54} \ldots\ldots x_{104} \times 2^{-52} \\ s_1^{(1)} = 1 + \left(x_{105} x_{106} \ldots\ldots x_{156} + x_{157} x_{158} \ldots\ldots x_{208} \right) \times 2^{-52} \\ s_2^{(1)} = 1 + \left(x_{157} x_{158} \ldots\ldots x_{208} + x_{209} x_{210} \ldots\ldots x_{260} \right) \times 2^{-52} \end{cases} \quad (19)$$

$$\begin{cases} a_0^{(2)} = x_1 x_2 \ldots\ldots x_{52} \times 2^{-52} \\ b_0^{(2)} = x_{53} x_{54} \ldots\ldots x_{104} \times 2^{-52} \\ c_0^{(2)} = x_{105} x_{106} \ldots\ldots x_{156} \times 2^{-52} \\ w_0^{(2)} = x_{157} x_{158} \ldots\ldots x_{208} \times 2^{-52} \\ \gamma^{(2)} = -0.3 + x_{209} x_{210} \ldots\ldots x_{260} \times 2^{-52} \end{cases} \quad (20)$$

Table 1 Correlation coefficient of plain and cipher video frames in the horizontal, vertical and diagonal directions

Size of the video frame	Type of the image	Name of the file	Correlation coefficient		
			Horizontal	Vertical	Diagonal
256 × 256	Plain video frame	5.1.08	0.9398	0.9024	0.9036
		5.1.09	0.8660	0.9051	0.8218
		5.1.10	0.9386	0.9575	0.8922
	Cipher video frame	5.1.08	0.0015	−0.0008	−0.0020
		5.1.09	−0.0015	−0.0009	−0.0039
		5.1.10	0.0043	0.0027	−0.47
512 × 512	Plain video frame	5.2.10	0.8921	0.9367	0.8581
		5.2.07	0.8660	0.9009	0.8029
		5.2.08	0.9297	0.9402	0.8973
	Cipher video frame	5.2.10	0.0018	−0.0026	−0.0052
		5.2.07	−0.0006	0.0019	0.0023
		5.2.08	0.0011	−0.0018	−0.0038
1024 × 1024	Plain video frame	5.3.02	0.9821	0.9776	0.9668
		5.3.03	0.9023	0.9105	0.8591
		7.2.02	0.9476	0.9647	0.9449
	Cipher video frame	5.3.02	0.0013	−0.0021	−0.0005
		5.3.03	0.0017	0.0005	−0.0005
		7.2.02	0.0007	0.0013	0.0008

A correlation coefficient between plain and cipher video frames is presented in Tab. 1. This correlation coefficient is presented in the horizontal, vertical, and diagonal orientations. Information regarding the various image sizes and the types that correspond to them is included in the table. In this particular context, there are two different types of video frames that are being discussed. The plain video frame and the cipher video frame are the two categories of video frames that make up the video frame. The plain video frame is obtained from the video that has not been encrypted, whereas the cipher video frame is obtained from the video that has been encrypted.

The correlation coefficient is a way of representing the degree of the non-linear similarity between the various components of the image. This configuration allows you to observe components that are horizontal, vertical, and diagonal in orientation. Strong positive relationships are indicated by correlation coefficients that are close to 1, while strong negative relationships are indicated by correlation coefficients that are close to –1. Fig. 4 depicts the correlation coefficient of the video frame of size 256 × 256. Let us discuss the correlation coefficient of different video frame.

In case of plain video frames with size 256 × 256, the correlation coefficient along the horizontal direction is 0.9398. This is an indication that the positive correlation is strong. In the vertical and diagonal directions, the value of the correlation is 0.9024 and 0.9036 respectively. This also indicates that the positive relationship is stronger.

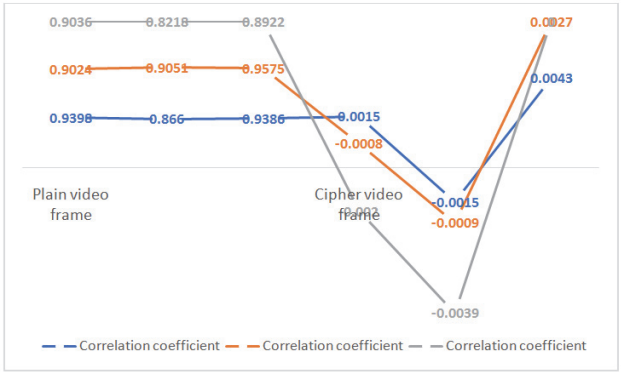

Figure 4 Correlation coefficient of the video frame of size 256 × 256

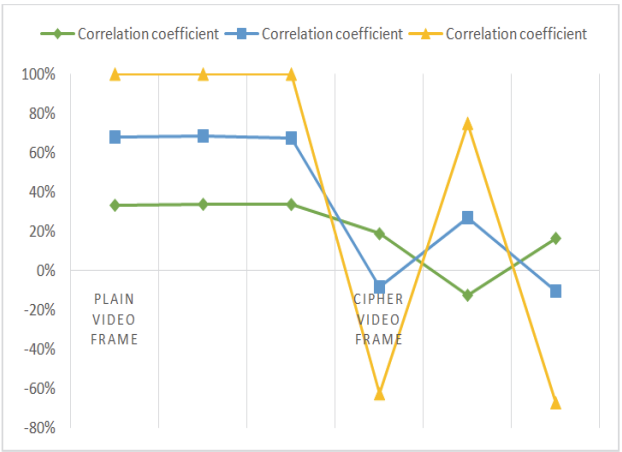

Figure 5 correlation coefficient of the video frame of size 512 × 512

Fig. 5 denotes the correlation coefficient of the video frame of size 512 × 512. In case of the cipher video frame,

the correlation coefficient along the horizontal direction is 0.0015. This means, they exist very weak positive correlation. Similarly, in the vertical and diagonal directions the correlation values are −0.0008 and −0.0020 respectively. All these are the indication that there is weak negative correlation. A strong positive correlation in the plain video frames denotes that the relationship among the pixels is strong and a weak correlation among the cipher video frames indicates that there is not much strong relationship among the pixels.

Tab. 2 provides the data related to the information entropy of video frames of different sizes. The information entropy is the randomness in the distribution of the pixels. The higher the value of the entropy, the higher the uncertainty and difficulty in the prediction. The plain video frames have higher entropy values than the cipher video frames. For example, the entropy for the files 5.2.07, 5.2.08 and 5.2.09 of the plain video frames are 7.2011, 6.9941 and 5.7057 respectively. The entropy is intentionally reduced in case of the cipher video frames in order to hide the information. Entropy plays a vital role in the encryption and the compression techniques.

Table 2 Information entropy of video frames of different sizes.

Size of the video frame	Name of the file	Value of the information entropy	
		Plain video frame	Cipher video frame
512 × 512	5.2.07	7.2011	7.9993
	5.2.08	6.9941	7.9993
	5.2.09	5.7057	7.9994
	7.1.02	6.0275	7.9995
	7.1.03	4.0056	7.9993
	7.1.02	5.4958	7.9993
	7.1.05	6.1075	7.9994
	7.1.06	6.5634	7.9995
	7.1.07	6.6954	7.9994
	7.1.08	5.9917	7.9994
	7.1.09	5.05345	7.9994
	Boat.512	7.1915	7.9993
	Elaine.512	7.5061	7.9995
	Gray21.512	4.3924	7.9995
	Numbers.512	7.7293	7.9993
	Ruler.512	0.6000	7.9992
1024 × 1024	5.3.02	7.5238	7.9999
	5.3.03	6.8304	7.9999
	7.2.02	5.6416	7.9999
	Test.1k	4.4078	7.9999

Table 3 PSNR of the encrypted and the decrypted video

S. No	Size of the block	PSNR of the decrypted video / %	PSNR of the encrypted video / %
1.	1 × 1	37.9257	12.6589
2.	4 × 4	37.1059	13.9798
3.	8 × 8	35.1102	14.1223
4.	16 × 16	33.9821	14.9878

Tab. 3 gives the PSNR of the encrypted and the decrypted video and Fig. 6 depicts the graphical representation of the PSNR of the encrypted and the decrypted data. For the block size of 1 × 1, the percentage of decrypted video is 37.9257 and that of the encrypted video is 12.6589. The PSNR of decrypted and the encrypted video for the block size of 4 × 4 are 37.1059% and 13.9798% respectively.

The PSNR of the decrypted video with the block sizes 8 × 8 and 16 × 16 are 35.1102% and 33.9821% respectively. Similarly, the percentages of the encrypted video are 14.1223% and 14.9878% respectively for the block sizes of 8 × 8 and 16 × 16.

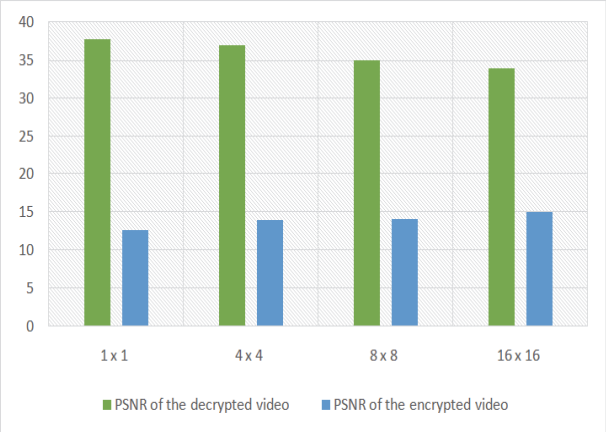

Figure 6 Graphical representation of the PSNR of encrypted and decrypted video

Tab. 4 gives the average coefficient of pixel of the encrypted frame for the first frame of telemedicine sessions video. The video taken here is the telemedicine sessions video. The average coefficients of the first frame of the selected video have been analyzed. The granularity in the sizes of the images is represented by the block sizes.

The average coefficient is the value in average related to pixels in the encrypted frame. The coefficients have been calculated separately for the horizontal, vertical and the diagonal directions.

The average coefficients of pixels of the encrypted frame for the 16 × 16 block are 0.0018%, 0.0044% and 0.0032% in the horizontal, vertical and the diagonal directions respectively. In case of 8 × 8 block, the average coefficient in the horizontal direction is 0.0016%, in the vertical direction is 0.0038% and in the diagonal direction is 0.0020%. the horizontal direction produced an average pixel coefficient value of 0.0024% and 0.0008% for the block sizes of 4 × 4 and 1 × 1 respectively. The vertical and the diagonal directions have average pixel coefficients of 0.0024% and 0.0008% for the 4 × 4 block and 0.0017% and 0.0013% for the 1 × 1 block respectively.

Table 4 Average coefficient of pixel of the encrypted frame for the first frame of telemedicine sessions video

S. No	Actual video	Size of the block	Avg coefficient of pixel of encrypted frame / %		
			Horizontal	Vertical	Diagonal
1.	Frame 1 of telemedicine sessions video	16 × 16	0.0018	0.0044	0.0032
2.		8 × 8	0.0016	0.0038	0.0020
3.		4 × 4	0.0008	0.0024	0.0017
4.		1 × 1	0.0004	0.0008	0.0013

Tab. 5 gives the PSNR of middle crop, side crop and salt and pepper noise and Fig. 7 depicts the graphical representation of the PSNR of middle crop, side crop and salt and pepper noise. Comparing the peak signal to the noise ratio is one method that can be utilized to evaluate the quality of the video that has been reconstructed. It denotes the variations between the original frame and the resultant frame. The image quality is observed to better when the value of the PSNR is high.

Table 5 PSNR of middle crop, side crop and salt and pepper noise

Video	Middle crop PSNR in dB	Side crop PSNR in dB	Salt and pepper noise PSNR in dB
Telemedicine sessions	27.45	24.78	30.45
Patient consultations	29.47	25.16	28.61
Surgical procedures	27.98	26.99	27.99

The analysis in the table is related to the PSNR at the central part of the frame, side regions of the frame and the frames that have salt and pepper noise. The PSNR can be used in the applications like video compression, noise removal and quality assessment. A higher PSNR value also tells that there is less distortion.

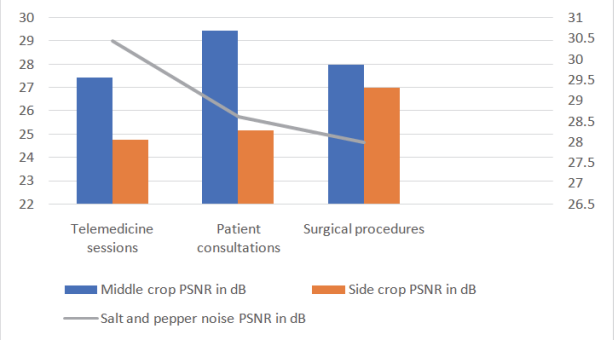

Figure 7 Graphical representation of PSNR of middle crop, side crop and salt and pepper noise

The PSNR of the middle and the side part of the video for the Telemedicine sessions video are 27.45 dB, 24.78 dB and 30.45 dB respectively. Similarly, the middle crop PSNR of the patient consultations, surgical procedures and the balcony are 29.47 dB, 27.98 dB and 28.79 dB respectively. The side crop PSNR of the patient consultations, surgical procedures and the balcony are 25.16 dB, 26.99 dB and 26.88 dB respectively. The salt and pepper noise PSNR of the patient consultations, surgical procedures and the balcony are 28.61 dB, 27.99 dB and 29.78 dB respectively.

Tab. 6 gives the mean absolute error, rate of pixel variation and average intensity change for different frames of the telemedicine sessions video. The mean absolute error of the first frame of the telemedicine sessions video is 87.6952. The rate of pixel variation for the frame 2 of the telemedicine sessions video is 99.4548. the average intensity variation of the fourth frame of the telemedicine sessions video is 28.99.

Table 6 Mean absolute error, Rate of pixel variation and average intensity change for different frames of the telemedicine sessions video

Video	Mean absolute error	Rate of pixel variation	Average intensity change
Frame 1 telemedicine sessions	87.6952	99.6327	35.01
Frame 2 telemedicine sessions	86.0917	99.4548	30.47
Frame 4 telemedicine sessions	86.3743	99.7410	28.99

Tab. 7 gives the encryption time of telemedicine sessions, patient consultations and Surgical procedures video. Fig. 8 gives the graphical representation of Encryption time of telemedicine sessions, patient consultations and Surgical procedures videos. The values represent the encryption time for every block size. When the block size is 1 × 1, the encryption times of the telemedicine sessions, patient consultations and surgical procedures videos are 0.9875, 0.8799 and 0.6199 respectively. When dealing with smaller block sizes, the encryption process will perform more efficiently in terms of time. Utilizing a block size that is larger will result in an encryption process that takes more time.

Table 7 Encryption time of different video frames of distinct block sizes

Actual video	Encryption time			
	1 × 1 block	4 × 4 block	8 × 8 block	16 × 16 block
Telemedicine sessions	0.9875	0.8145	0.7155	0.6016
Patient consultations	0.8799	0.7472	0.6416	0.5749
Surgical procedures	0.6199	0.4013	0.2216	0.1978

The method balances encryption speed and robustness by using an optimized S-box with high non-linearity for security and efficient lookup operations for speed. This ensures strong resistance to attacks without compromising processing time, making it suitable for real-time and resource-constrained applications.

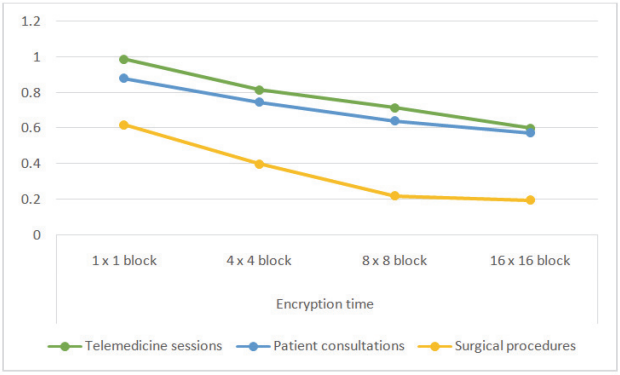

Figure 8 Graphical representation of Encryption time of telemedicine sessions, Patient consultations and Surgical procedures videos

The performance of the proposed work is compared with the state of art methodologies in terms of Peak Signal to Noise Ratio (PSNR), Mean Absolute Error (MAE), Encryption and Decryption Time. The observed values are tabulated in Tab. 8.

Table 8 Performance comparison

State of Art Methods	PSNR	MAE	Encryption Time	Decryption Time
X. Zhou [20]	35.21	33.21	0.3415	0.3315
A. Al-Hyari [21]	33.48	31.47	0.3174	0.3156
Z. Zhang [22]	36.15	32.15	0.3384	0.3240
J. Shao [23]	30.04	30.11	0.2415	0.2348
J. Li [24]	30.43	30.06	0.2365	0.2262
Proposed Work	27.99	28.99	0.1978	0.1823

The optimized S-box in the proposed work enhances PSNR and reduces MAE by ensuring better non-linearity

and uniform transformations, preserving data fidelity during encryption and decryption. Its efficient design reduces computational complexity, enabling faster encryption and decryption processes.The computational complexity of the S-box depends on its size and lookup operations, typically O(1) for precomputed tables, ensuring fast substitutions. Resource requirements include memory for storing the S-box and minimal processing power, making it efficient for hardware and software implementations.

The security of an S-box is evaluated against attacks like linear and differential cryptanalysis, brute force, and algebraic attacks. High non-linearity and balanced input-output mappings enhance resistance to linear and differential cryptanalysis by minimizing predictable patterns. Avalanche properties ensure that small input changes result in significant output differences, thwarting brute force and statistical attacks. Algebraic immunity, achieved through complex Boolean functions, prevents attackers from solving equations representing the S-box. A well-designed S-box with these properties strengthens the overall cryptographic algorithm's security.

5 CONCLUSION

In this paper, an enhanced video encryption technique using the s box and the spatial filtering has been performed. The initial values of the state have been generated by using the binary sequence with the length of 260. This acts as the private key. This avoids the constant change in the private key and it also diminishes the need for the recreation of the chaotic sequence. After the conversion of the two-dimensional video into the three dimensional video, spatial permutation, plane level filtering and superposition were traced at the video frames which could be initiated at all the axes (x, y and z). Furthermore, substitution boxes are used for the enhancement of the security in the video encryption. The diffusion of pixels into the neighboring pixels has been avoided by performing random swapping operation after which the encrypted video can be obtained. The proposed technique has proven to provide better parametrical results in terms of the efficiency and security. The potential challenges include managing computational overhead for high-resolution videos and ensuring real-time performance in resource-constrained environments. Future scope involves optimizing the method for larger datasets, exploring adaptive S-box designs, and integrating machine learning for dynamic encryption strategies.

7 REFERENCES

[1] Kumar, V., Pathak, V., Badal, N., Pandey, P. S., Mishra, R., & Gupta, S. K. (2022). Complex entropy based encryption and decryption technique for securing medical images. *Multimedia Tools and Applications*, *81*(26), 37441-37459. https://doi.org/10.1007/s11042-021-11822-3

[2] Boyd, D. & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662-679. https://doi.org/10.1080/1369118X.2012.678878

[3] Robson, M. (2014). Signals in the sea: The value of Ultra intelligence in the Mediterranean in World War II. *Journal of Intelligence History*, *13*(2), 176-188. https://doi.org/10.1080/16161262.2014.925334

[4] Kavanaugh, A., Fox, E. A., Sheetz, S., Yang, S., Li, L. T., Whalen, T., Shoemaker, D., Natsev, P., & Xie, L. (2011). Social media use by government: From the routine to the critical. *Proceedings of the 12th annual international digital government research conference: Digital government innovation in challenging times*, 121-130. https://doi.org/10.1145/2037556.2037578

[5] Prabakaran, D. & H. Sathyapriya. (2019). A review on methodologies and performance analysis of device identity masking techniques. *Journal of Scientific & Technology Research*, *8*(12), 2018-2022.

[6] Paar, C. & Pelzl, J. (2009). Understanding cryptography: A textbook for students and practitioners. *Springer Science & Business Media*. https://doi.org/10.1007/978-3-642-04101-3

[7] Roberts, D. L. (2019). Republic of numbers: Unexpected stories of mathematical Americans through history. *Johns Hopkins University Press*. https://doi.org/10.1353/book.73771

[8] Norouzi, B., Mirzakuchaki, S., Seyedzadeh, S. M., & Mosavi, M. R. (2014). A simple, sensitive and secure image encryption algorithm based on hyper-chaotic system with only one round diffusion process. *Multimedia Tools and Applications*, *71*, 1469-1497. https://doi.org/10.1007/s11042-013-1353-6

[9] Buchanan, W. J., Li, S., & Asif, R. (2017). Lightweight cryptography methods. *Journal of Cyber Security Technology*, *1*(3-4), 187-201. https://doi.org/10.1080/23742917.2017.1405600

[10] Mushtaq, M. F., Jamel, S., Disina, A. H., Pindar, Z. A., Shakir, N. S. A., & Deris, M. M. (2017). A survey on the cryptographic encryption algorithms. *International Journal of Advanced Computer Science and Applications*, *8*(11). https://doi.org/10.14569/IJACSA.2017.081142

[11] Murari T, V., Ravishankar, K. C., & Raghu, M. E. (2024). Selective encryption of video frames using the one-time random key algorithm and permutation techniques for secure transmission over the content delivery network. *Multimedia Tools and Applications*, 1-40. https://doi.org/10.1007/s11042-024-14913-z

[12] Liu, J., Li, Y., Han, G., & Sun, N. (2023). Visual video evaluation association modeling based on chaotic pseudo-random multi-layer compressed sensing for visual privacy-protected keyframe extraction. *Journal of Visual Communication and Image Representation*, *90*, 103691. https://doi.org/10.1016/j.jvcir.2022.103691

[13] Noor, N. S., Hammood, D. A., Al-Naji, A., & Chahl, J. (2022). A fast text-to-image encryption-decryption algorithm for secure network communication. *Computers*, *11*(3), 39. https://doi.org/10.3390/computers11030039

[14] Bouaafia, S., Khemiri, R., Messaoud, S., Ben Ahmed, O., & Sayadi, F. E. (2022). Deep learning-based video quality enhancement for the new versatile video coding. *Neural Computing and Applications*, *34*(17), 14135-14149. https://doi.org/10.1007/s00521-022-07359-5

[15] Sha, K., Wei, W., Yang, T. A., Wang, Z., & Shi, W. (2018). On security challenges and open issues in Internet of Things. *Future Generation Computer Systems*, *83*, 326-337. https://doi.org/10.1016/j.future.2018.01.059

[16] Abomhara, M., Khalifa, O. O., Zaidan, A. A., Zaidan, B. B., Zakaria, O., & Gani, A. (2011). An experiment of scalable video security solution using H.264/AVC and advanced encryption standard (AES): Selective cryptography. *International Journal of the Physical Sciences*, *6*(16), 4053-4063.

[17] Abduljabbar, Z. A., Abduljaleel, I. Q., Ma, J., Al Sibahee, M. A., Nyangaresi, V. O., Honi, D. G., Abdulsada, A. I., & Jiao, X. (2022). Provably secure and fast color image encryption algorithm based on s-boxes and hyperchaotic map. *IEEE Access*, *10*, 26257-26270. https://doi.org/10.1109/ACCESS.2022.3156750

[18] Wen, H., Lin, Y., Kang, S., Zhang, X., & Zou, K. (2024). Secure image encryption algorithm using chaos-based block permutation and weighted bit planes chain diffusion. *iScience*, *27*(1), 105855. https://doi.org/10.1016/j.isci.2023.105855

[19] Ramachandran, S. & Prabakaran, D. (2024). A Novel Deep Reinforcement Learning (DRL) Method for Detecting Evasion Attack in Iot Environment. *2024 International Conference on Integrated Circuits and Communication Systems (ICICACS)*, 1-6. https://doi.org/10.1109/ICICACS60521.2024.10498632

[20] Zhou, X., Wang, H., Li, K., Tang, L., Mo, N., & Jin, Y. (2024). A Video Streaming Encryption Method and Experimental System Based on Reconfigurable Quaternary Logic Operators. *IEEE Access*, *12*, 25034-25051. https://doi.org/10.1109/ACCESS.2024.3365523

[21] Al-Hyari, A., Obimbo, C., Abu-Faraj M. M., & Al-Taharwa, I. (2024). Generating Powerful Encryption Keys for Image Cryptography with Chaotic Maps by Incorporating Collatz Conjecture. *IEEE Access*, *12*, 4825-4844. https://doi.org/10.1109/ACCESS.2024.3349470

[22] Zhang, Z., Tang, J., Zhang, F., Huang, T., & Lu, M. (2024). Medical Image Encryption Based on Josephus Scrambling and Dynamic Cross-Diffusion for Patient Privacy Security. *IEEE Transactions on Circuits and Systems for Video Technology*, *34*(10), 9250-9263. https://doi.org/10.1109/TCSVT.2024.3394951

[23] Shao, J., Bai, E., Jiang, X., & Wu, Y. (2024). Multi-View Light Field Images Compression and Encryption Using Enhanced 3D Chaotic System and Pixel-Bit-Scrambling. *IEEE Access*, *12*, 156471-156491. https://doi.org/10.1109/ACCESS.2024.3481230

[24] Li, J. & Hou, X. (2024). The Design of an Adaptive Enhanced AMP-Based Image Block Compressed Sensing and its Application to Image Encryption. *IEEE Transactions on Circuits and Systems for Video Technology*, *34*(11), 11128-11141. https://doi.org/10.1109/TCSVT.2024.3417287

Contact information:

R. SAKTHEESWARI
Department of IT,
Sri Venkateswara College of Engineering,
Chennai, Tamil Nadu, India - 602117

C. YAASHUWANTH
(Corresponding author)
Department of IT,
Sri Venkateswara College of Engineering,
Chennai, Tamil Nadu, India - 602117
E-mail: yaash_it@rediffmail.com

K. PRATHIBANANDHI
Department of EEE,
Sri Sairam Engineering College,
Chennai, Tamilnadu, India - 602109

S. RAMESH
Department of CSE,
A.K.T. Memorial College of Engineering and Technology,
Kallakurichy, Tamilnadu, India - 606213