Journal of Autonomous Intelligence (2023) Volume 6 Issue 2
doi: 10.32629/jai.v6i2.665

REVIEW ARTICLE

A systematic review on video encryption algorithms: A future research

Avnish Kanungo, Ayushi Srivastava, Saniya Anklesaria, Prathamesh Churi*

Department of Computer Engineering, Mukesh Patel School of Technology Management & Engineering, NMIMS University, Mumbai 400056, India

*** Corresponding author:** Prathamesh Churi, prathamesh.churi@gmail.com

ABSTRACT

Video encryption is widely used in many real-time applications today. Despite numerous video encryption techniques that are available today, the challenges such as time and space complexity, real time latency, scalability and vulnerability towards few attacks still exist in the research domain, however few algorithms have achieved acceptable computational complexity, but in such cases vulnerability to certain attacks (differential, statistical, plain text attack, and cipher text attacks) are still a threat to secure video transmission. To the best of our knowledge, a comprehensive but detailed systematic literature review on video encryption is needed for the researchers in the scientific community. This paper, therefore, presents a systematic literature review of 30 scientific documents extracted from platforms like scopus and web of science. The paper, comprehensively addresses, various techniques of video encryption and encoding, different evaluation parameters to testify the performance of the algorithms and discusses the challenges of the existing video encryption algorithms. After careful investigation, it has been observed the approaches which involve encryption of the video data post the same has been encoded is the most efficient and scalable approach towards video encryption. In addition, it also implies that in near future, the proposed algorithms, must be evaluated based on the various categorization of parameters illustrated in this paper.

Keywords : video encryption; chaos; frames; encoding; parameters; algorithms

ARTICLE INFO

Received: 26 May 2023
Accepted: 16 June 2023
Available online: 1 August 2023

1. Introduction

Video encryption is widely used in many real-time applications today. Drones, or unmanned aerial vehicles, are utilised in numerous applications where it is necessary to record or broadcast video. Drones store video data and transmit them to a control centre where additional analysis can be performed on the data in the Intelligent Transportation Systems (ITS) and public safety applications. Various clientele, including police enforcement and emergency services employees, share these films. In such circumstances, the video could contain the faces of bystanders as well as other confidential information that might raise privacy issues and therefore need to be encrypted[1]. Another application is for copyrights as the amount of digital information that is shared and distributed across various online platforms, particularly video, has significantly expanded in the modern internet era[2]. Other applications include military, medical, drones, satellite, space use cases and drone applications[1,2].

Video, or moving pictures, is a series of still images (called frames) that are played back in fast succession. It is becoming increasingly simple to intercept and steal sensitive information (including text,

photos, audio, video, etc.) being carried via the internet since it is an unsecured medium[3,4]. The security of multimedia files is of essential importance because of the prevalence of confidential data inside them. Security measures for data transmission via the internet have included encryption, authentication, and digital signatures. Several cryptographic protocols have been suggested for encrypting videos, including the Advance Encryption Standard (AES), RC4, Simplified Data Encryption Standard (SDES), Modified Advance Encryption Standard (MAES), etc.[3,5,6].

Over the years, video encryption has improved drastically trying to overcome the research gaps. Current video encryption techniques solve the problem of dealing with MPEG as well as there have been improvements in the H.264 while some do not need any additional hardware anymore[3]. Some limitations that were observed after the comprehensive study of these papers included susceptibility to geometrical attacks such as rotation and flipping, limited data gathering from distorted decoded video and dealing with time complexity and spatial complexity. Multiple methods have been attempted throughout the years to improve visual security such as integration of amplitude encoding, integrating CCRM camera, and improving visual scrambling.

The main contributions of this paper are:

- Throws light on the different approaches and techniques used in video encryption like AES, SDES, MAES (complete this sentence).
- Categorise the evaluation parameters of existing video encryption algorithms such as Security Metrics, computational cost, and evaluate how do they differ from evaluation parameters of image encryption algorithms.
- Identify the challenges and limitations of existing video encryption algorithms.

This research discusses a broad variety of approaches to key generation, from chaotic system and cellular automaton-based pseudo-random number generators to the use of various encryption algorithms (such as AES, DES, Rabbit, RSA, Motion Vector and others) to ensure the safety of stored and transmitted data. Many of the above-mentioned techniques, key generation procedures, and encryption algorithms are used in the methodology of the research articles under review. The strongest discussion in the study is that we have come up with our own categorisation for the evaluation parameters and also mentioned how the papers under review connect from the technical to real world.

Some inputs that this research give to the encryption/security researchers, readers, PhD scholars, media servers and practitioners involve identifying which technique of video encryption is the most suitable along with what parameters are the best to test the method. Some areas that they can work on can include minimizing the encryption time, dealing with big data issues, dealing with the issues of AES and DES amongst the others which are discussed further in the Literature Review section.

The paper is divided in the following sections. Section 2 talks about the search criteria and research questions, followed by section 3 that covers the first research question: approaches and techniques in video encryption. Section 4 discusses the second research question: evaluation parameters of video encryption algorithms. Further section 5 describes the challenges and limitations of video encryption Algorithms which forms the third research question. Finally, section 6 includes the discussions and conclusions of the paper.

2. Search criteria and research questions

For the systematic review carried out in this paper, we began by searching documents using various strings such as, “Video Encryption Techniques”, “Evaluation of Video Encryption Algorithm”, “Video Encryption + Chaotic Maps”, “Video Encryption + Selective”, “Video Encryption + DNA Coding” etc. and shortlisted the most relevant papers. The papers were selected based on the publications (majority of the papers were Springer, IEEE, Elsevier, Inderscience etc.) and indexed papers from Scopus. A total of 57 documents were retrieved out of which a few were removed as they were not relevant to our study. The papers that focused

on data hiding were removed. We also focused on the encryption algorithms proposed recently so papers prior 2017 were also removed. Total 30 papers were taken into consideration which are distributed over five years from the year 2017 to the year 2022. **Figure 1** describes the keywords, inclusion and exclusion criteria of the systematic review. The papers cover topics such as selective video encryption, video encryption using chaotic maps and the performance evaluation of these algorithms. Based on the previous work done in this field, we came up with three research questions and identified the main points of our review.

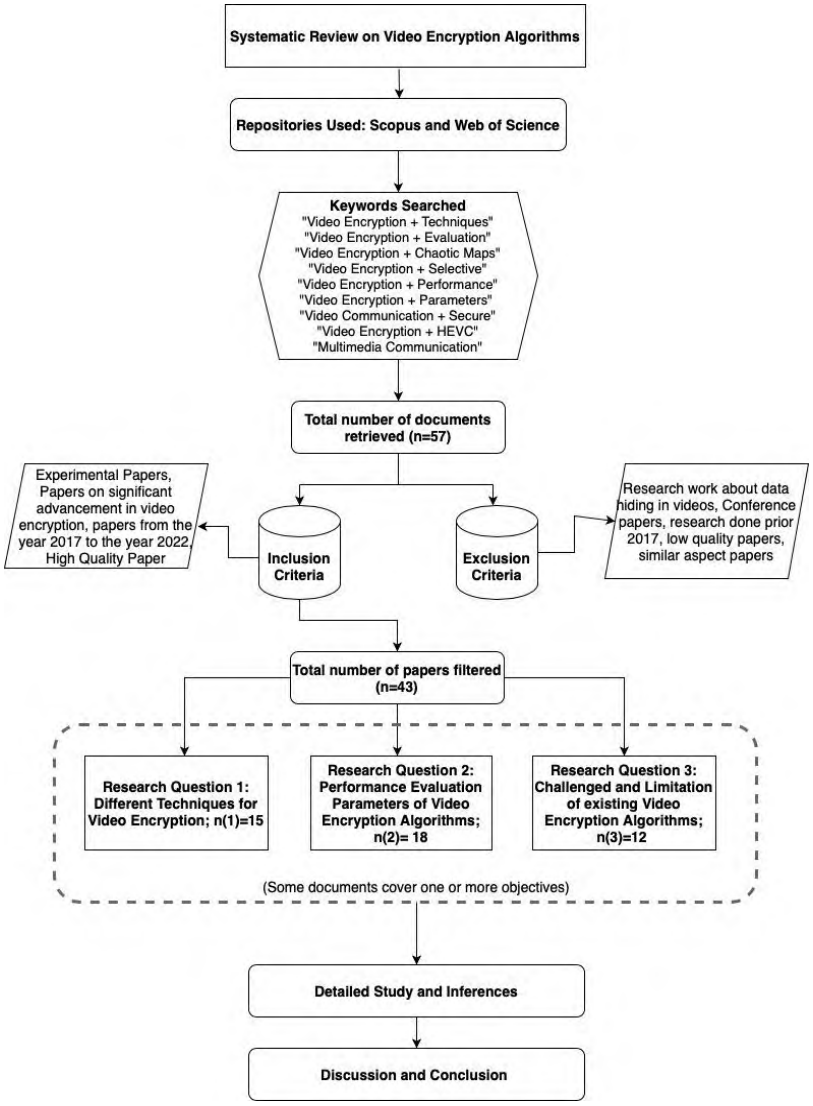

Figure 1. Search criteria, inclusion/exclusion criteria of our systematic review.

Our systematic review on video encryption Techniques aims to answer the following research questions:

RQ1: What are the different approaches and techniques used in video encryption?
RQ2: What are the different categories of evaluation parameters of existing video encryption algorithms?
RQ3: What are the challenges and limitations of existing video encryption algorithms?

In the following sections, we seek to answer the research questions after analysing the relevant papers.

3. Approaches and techniques in video encryption (RQ1)

This section covers the various approaches that one can employ to a video encryption algorithm. It also talks about the different techniques that are used in video encoding and video encryption algorithms. **Figure 2** shows the overview of these approaches and techniques.

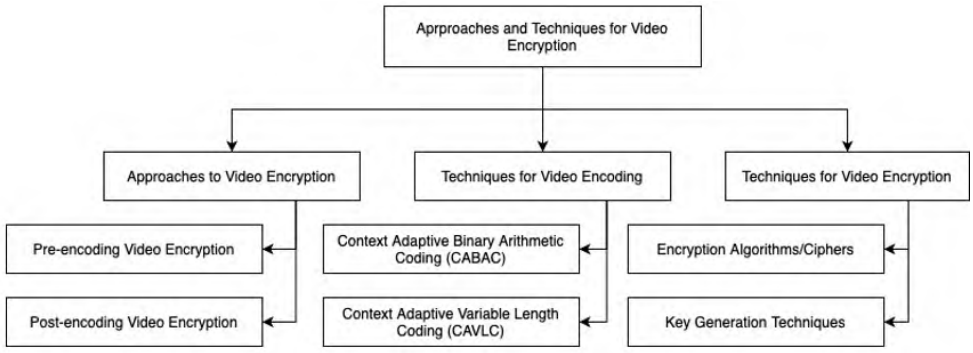

Figure 2. Approaches and techniques in video encryption.

3.1. Approaches to video encryption

Video encryption approaches are generally classified into selective and naive encryption. Naïve encryption, refers to the approach that encrypts all the frames, parameters or data comprised in a video stream. This kind of encryption is expensive in terms of both time and space, and requires a substantial amount of computational power. The advantage of this is that as all relevant data has been encrypted, the level of security is quite high depending on the techniques used. But the same level of security can also be achieved by encryption of strategically chosen frames or information bits, which is where selective encryption comes in the research of Abomhara et al.[3]. In this approach, using evaluation techniques we select frames or information bits that have the highest impact on processing of the video and those are the parameters that are chosen for encryption, as fewer amount of data needs to be encrypted, this provides for a faster and computationally less complex approach to encryption, due to which majority of the modern video encryption methodologies employ this approach[4]. The conversion of a digital video's format from one standard to another, usually for compatibility reasons, is known as video encoding. Encoding also helps to decrease the overall size of the video file via compression to facilitate the process of transmission. The process of encoding (at the source) and decoding (at the receiver) is done by a software known as a video codec. The Video codec on the encoder end compresses the video data to a standardized format and transmits this data as a bit stream, which when received by the decoder on the receiver's end is decompressed and is available for viewing in the original format. The Video codecs are available for different coding formats like MPEG, H.264/AVC, H.265/HEVC and many more. The most prevalent format is H.264/AVC with the newer version of the same H.265/HEVC, slowly catching up. The coding standards primarily associated with these formats are CABAC and CAVLC. Hence, the majority of the papers reviewed are based on these two formats.

Based on the process of encoding, encryption can take place before the encoding process or after the encryption process, on the basis of this we have defined the approaches to encryption as pre-encoding, encryption before the encoding takes place by the codec or post-encoding, where the process of encryption takes place after the encoding process or during the encoding process by augmenting the codecs to carry out the encryption process too. Under selective encryption on the basis of the sequence when encoding will occur the approaches can be further divided in the following.

3.1.1. Pre-encoding encryption techniques

The pre-encoding encryption processes contain processes where frame-based encryption happens, i.e., in this type of encryption, we take the actual video stream, break it down into individual frames and apply encryption on the selected frames of the video. The distortion of the images is done on the basis of processes of confusion, diffusion or both applied via mathematical transforms. They have been so named, due to the fact that the encryption process takes place before the encoding process. **Figure 3** shows the process of applying encryption algorithm before the video is encoded.

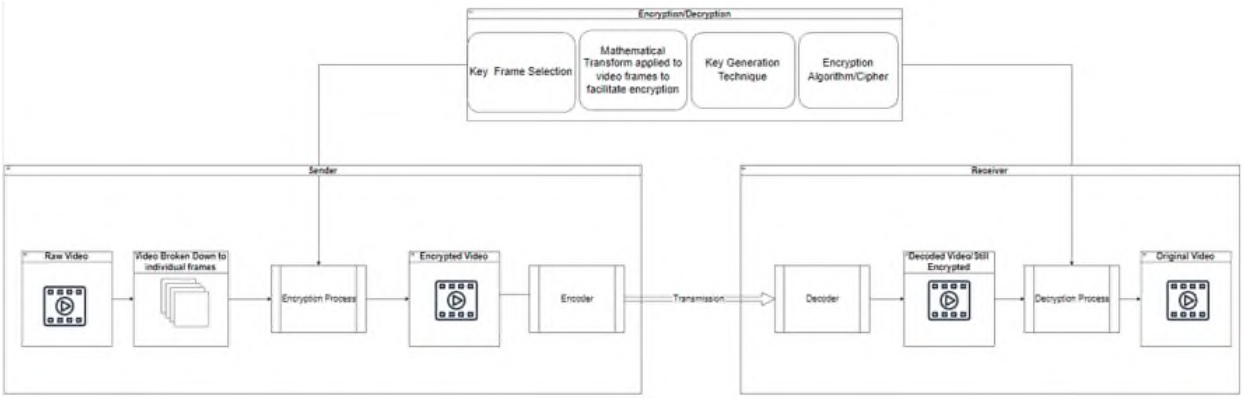

Figure 3. Pre-encoding encryption algorithm.

Transforms used for the process of encryption:

- **Jigsaw state transform**

The Jigsaw transform is a non-linear transposition function that alternates the adjacent blocks of an image. The unitary conservation of energy during the transformation process is one of the characteristics of the Jigsaw transform. The Jigsaw transform additionally offers an inverse function. The Jigsaw transform can appear in a 2D format or 3D format[4]. The Jigsaw transform basically jumbles the blocks for multiple iterations to create contorted iterations of the frame.

- **Fractional Fourier Transform**

The 2D-Fractional Fourier Transform is a linear transformation that rotates the input signal by any angle into a required phase as part of the transform space domains. It is regarded as a reduction of the traditional Fourier Transform (FT) technique. The suggested video cryptography methods use the Fractional Fourier Transform since it is widely used in the field of optics, namely optical statistics and signal processing applications, hence by extension it is indispensable in the field of image and video encryption[4]. The arbitrary phase, which is a component of the Fourier plane, functions as the ciphering process's secret key. The additional degree of freedom provided by the 2D-FrFT aids in increasing the size and complexity of the secret key, which enhances the security of the encryption further. The 2D-FrFT is thus used in a number of cryptographic methods for the aforementioned benefits.

- **SVD (Singular Value Decomposition)**

Given a matrix X of size m × n, the singular value decomposition of X results in three matrices (U, S, and V). U and V are orthogonal matrices with size m × m, whereas S is a diagonal matrix containing values called "the singular values"[5]. Equation below shows the equation for SVD:

$$X = U \times S \times Vt \text{ (Specify variables)}$$

The diagonal values of the S matrix are the values that are considered the most significant values and are considered for further encryption and transformation for data hiding[6]. The use of SVD is also prevalent in the watermarking techniques, where the S values from the SVD transform of the watermark is applied, added to the frame via a transform and used.

- **Sparse representation using over complete dictionary**

According to this model, a dictionary D can be utilized to sparsely depict a given signal as follows, ($D \in Rn \times k$) and it needs to be ensured that this dictionary is overcomplete ($k > n$) in nature which will allow us to harness this characteristic of the dictionary for our benefit. The general sparse representation of a signal vector (x) is given as,

$$x = D\alpha$$

where α is the sparsest possible representation of the original signal satisfying. Thus, the solution will be non-unique, and allows to absorb some error[7]. To get the sparse form (α) of a signal, we have to solve the following equation.

$$\alpha = arg\min_{\alpha} \|\alpha\|0 \ \ subject\ to\ \|x - D\alpha\|_2^2 \leq \epsilon$$

The dictionary in question is trained by using a dictionary training algorithm (e.g., matching pursuit). In order to reduce error, the trained dictionary is updated after each iteration, producing an optimized sparse vector, α.

- **Rubik's cube algorithm**

This algorithm is used to scramble the pixels of the original image. The Rubik's cube algorithm works on the idea of circularly shifting the image bits in the direction and for the number of iterations selected during the algorithm configuration[8]. Using two random secret keys, the bitwise XOR is applied into the odd rows and columns. Then, the bitwise XOR is also applied to even rows and columns using the flipped secret keys. These steps can be repeated till the number of iterations specified are reached. This can be visualized as the frame being wrapped around a Rubik's cube and the cube being shuffled as per the number of iterations defined.

For the Pre-encoding encryption Techniques, for a number of approaches, special techniques for frame selection are also employed, which are discussed below:

- **Linear Discriminant Analysis (LDA) and Convolutional Neural Network (CNN)**

The method mentioned has two main parts. In the first part, frame-level video labelling is carried out utilizing the Linear Discriminant Analysis (LDA). Two CNNs are utilized to extract spatial and temporal information, and LDA is then used to find a projection that optimizes distance between distinct action videos while minimizing distance between the same action videos. For this, we use a dataset of human action videos where the action labels are given. A frame's distance from the class mean is utilized as a uniqueness metric after projection onto the LDA space. The likelihood that a frame is a key frame increases with uniqueness value. In the second step, a two stream CNN is trained to regress video frames on the uniqueness score obtained from the LDA projection of the first step. Once the CNN is trained, it is ready to output the uniqueness score of frames in a video and the frames corresponding to the peaks are selected as key frames[9].

- **Haar's technique**

The Haar cascade technique[10] is based on a classifier that is trained on a similar type of image data set. It is then used to identify the important areas for encryption in a frame extracted from a video. This is applied in a naive manner to all the frames of the video. The process includes integral feature selections, such as edge feature, line feature etc., which will help to uniquely identify a type of image. Then on the basis of these features we apply feature selection on the basis of boosting method, in which we iteratively apply weak classifiers and incrementally increase the weights of incorrectly classified weights to make sure that they are differentiated in the next round. This will allow classification of the features and allow us to choose the features that need to be used. This is extended to multiple object detection using CNNs, which will provide multiple objects of importance in the frame. On the basis of the examination of these features and objects of importance, we will then make the selection of the frames for encryption.

- **Change detection mechanism**

A scene change detection mechanism is used to identify significant frames, for the same adjacent frames are compared to one another and identical frames are grouped together. The value of the frame difference will decide whether the frame will be considered as the part of the same group or different group[6]. If difference is large, then it will be considered as part of a different group. The threshold will be used as the decision parameter; if the value of the frame difference is greater than the threshold, the frame will be included in the

next group. A temporal sampling is performed that enhances the process of frame selection. The videos with a larger number of scene changes will have a bigger number of frames selected as key frames.

3.1.2. Post-encoding encryption technique

The post-encoding encryption, as the name suggests is the encryption of the data generated by the video encoding mechanism, which happens after the encoding process has been carried out. This is generally performed on the basis of the NALU bits that are generated by the encoding mechanism, be it CABAC or CAVLC. These include the Motion vectors, DCT coefficients and their signing bit. Once the syntax to be encrypted are finalized, they are encrypted on the basis of a combination of key generation technique and encryption algorithm. (e.g., AES or AES version of stream cipher). **Figure 4** shows the process of encrypting the video after it has been encoded.

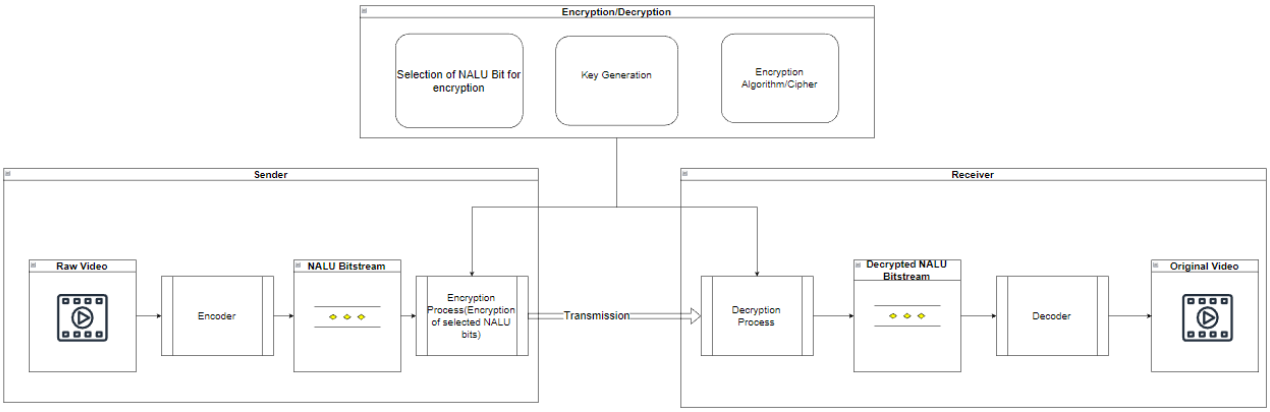

Figure 4. Post-encoding encryption algorithm.

NALU Bits used for encryption (choosing technique based on spatial influence):

- **Motion Vector (MVx, MVy)**

HEVC defines a signed 16-bit range for both horizontal and vertical motion vectors i.e., along the X and Y axis, hence MVx and MVy vectors. It is a two-dimensional vector used for inter-prediction that provides an offset from the coordinates in the decoded frame to the coordinates in a reference frame of a video[11]. Due to its importance in inter prediction, the motion vector and the signs of its individual bit coeffecients are considered as one of the parameters which, if encoded, will lead to the most distortion of data[12].

- **Transform Coefficients (TC)**

Image encoding techniques utilize mathematical transforms applied to the frame data. This effectively allows the data to be stored in a smaller amount of space, essentially helping in compression of data, the transform primarily used for the same is the discrete cosine transform (DCT), and both the values and the signs of the output bits are considered as parameters for encryption, as both are required to recreate the frame during the decoding process[11,13].

- **Start code**

Each top-level (Network Abstraction Layer) unit of the bitstream of the general standard video codecs (i.e., H.264/AVC, HEVC, and IVC) contains the start code as a prefix, which is a strong separator between the NAL units. Using this, a parser attempting to locate the start code in a bitstream can quickly split the bitstream into NAL units, without having to parse every syntax element of the bitstream. As the parsers an integral component of the decoder, the start codes form a very important part of the video decoding process[14]. Therefore, encrypting them will cause the decoder to not differentiate between the NAL units, leading to significant distortion of the recreated frame and over all videos.

- **Inter-prediction mode parameters (IPM chroma, IPM Luma)**

Inter prediction creates a prediction model from one or more previously encoded video frames[13]. The model is formed by shifting samples in the reference frame(s) (motion compensated prediction). The luminance component of the IPM model is called IPM luma and is used for predicting the brightness of the individual components, while the color component is the IPM chroma, which determines the overall coloration of the frames[12]. These parameters, if encoded, can lead to significant distortion in the video quality if the actual values are not used.

- **Residual syntax elements (signs and coefficients)**

The residual syntax components are responsible for edge detection and differentiation of elements of the frame from the surroundings[12]. The signs and values of the residual components both are considered for encryption due to their high impact in the process of video recreation during decoding[15].

3.2. Techniques for video encoding

3.2.1. Video encoding

Video encoding is the process of compressing the video in representative bit stream which can then be read by the decoder and the actual video can be reproduced at the recipient end. It is primarily required as it optimizes transmission of the video over the channel that chosen for the same. The video encoding technique that is used is primarily responsible for the representative syntax bits that are used for the encryption processes. The video encoding method used generally has a much greater impact on the encryption process if the encryption happens post encryption[11]. As the papers discussed primarily are concerned with video compression standards, H.264/AVC and H.265/HEVC. We will look at the encoding mechanism employed for these two standards:

- **CABAC (Context Adaptive Binary Arithmetic Coding)**

Video encryption methods like H.264/MPEG-4 AVC and High Efficiency Video Coding (HEVC) use context-adaptive binary arithmetic coding (CABAC) which is a sort of entropy encoding[13]. CABAC provides a higher compression than most other entropy encoding algorithms used in video encoding that too in a lossless manner even though its application in video codecs leads to a lossy implementation, which is one of the unique components of this encoding technique which allows it to perform better than its predecessors and provide a more efficient encoding process.

Arithmetic coding forms the foundation of this technique, which is further tweaked so as to support the preconditions set forth as per the video encoding standards. The major components for the same are described below[11]:

- Binarization: the procedure is kept simple by encoding all non-binary values to Binary symbols, which in turn also extends this simplicity to probability models which are used for modelling the frequently used bits of any symbol.
- Context Modelling Selection: because coding modes are usually well correlated locally, the probability models are selected adaptively based on local context, allowing for much more accurate probability modelling. The context-model, which is usually used to encode the incoming data on the basis of statistical modelling of the previous data, are generally selected from a set of available models, on the basis of the statistical properties of the existing bits.
- Arithmetic Coding and Probability Update: the arithmetic coder component then encodes each bit as per the context model selected and outputs the coded value. The context model is then updated on the basis of the coded value.

The technique uses a multiplication-free range division by the use of quantized probability ranges and probability states.

CABAC has multiple probability modes for different contexts. It first converts all non-binary symbols to binary. Then, for each bit, the coder selects which probability model to use, then uses information from nearby elements to optimize the probability estimate[12]. Arithmetic coding is finally applied to compress the data.

- **CAVLC (Context Adaptive Variable Length Coding)**

The second most prevalent coding technique, primarily used in H.264/MPEG-4 AVC video encoding is called Context-adaptive variable-length coding (CAVLC). This too is a form of entropy encoding which is lossless in terms of compression. It is used to encode the residual, zig-zag order blocks of the transform coefficient. It is a predecessor to CABAC and requires less processing to decode, which is due to the fact that it does not compress the data that effectively. Consequently, this method of encoding is primarily absent from latest implementations of video encoding.

3.3. Techniques for video encryption

The encryption process is generally a two-part process in the video encryption space, this includes generation of the encryption key using a pseudo-random number generation process, this key is then combined with the encryption algorithm by the means of an XOR gate, additive OR gate, matrix multiplications or the key might be used as a seed for a cyclic encryption algorithm. Below you will find the list of encryption algorithms and key generation techniques that have been used in the recent papers, the algorithms and key generation techniques can be mixed and matched as per the required level of encryption and the use case.

3.3.1. Encryption algorithms/ciphers

AES (Advanced encryption standard): AES is one of the most well-known standard block ciphers that is classified as a symmetric cipher[8,16,17]. AES is based on a design principle known as a substitution–permutation network, and is efficient in both software and hardware[12,18]. The components of AES are application of round key via XOR, substitution based on a lookup table, cyclic shifting of last rows as per the number defined and linear transformation process as per a defined vector[19]. Multiple iterations of these processes are carried out as per the complexity requirement.

AES-CTR mode: this is a specific mode of AES, which can be used to generate pseudo random numbers, using its property of cyclicity[12,20].

PRESENT: it is a lightweight block cipher, that is classified as a symmetric cipher. Its operation can be described as follows, a block of size 64 bits is taken along with a key which can be either 80 bits or 128 bits. Its major component is a non-linear layer which comprises a 4-bit substitution box which was optimized to operate on hardware[5]. The bits with the key are input into the non-linear layer to encrypt the data. This cipher is designed to operate in situations where low power consumption and high chip efficiency is required.

Rabbit: it is a high-speed stream cipher that is classified as a symmetric cipher. It is a 128-bit key and a 64-bit initialization vector and was designed specifically to be optimized for software applications[21]. The core component of the cipher is a bitstream generator which encrypts 128 message bits per iteration. The cipher's strength rests on a strong mixing of its inner state between two consecutive iterations.

RSA: it is one of the most used public-key cryptosystems that is widely used for secure data transmission. The RSA algorithm involves four steps: key generation, key distribution, encryption, and decryption. The basic principle behind RSA is the observation that it is practical to find three very large positive integers e, d, and n, such that with modular exponentiation for all integers m (with $0 \leqslant m < n$).

ECC: it is another type of public-key cryptosystems based on the algebraic symmetry of elliptic curves over finite fields, which are generally constrained to primes. ECC is based on multiple iterations of a function

which finds corresponding points to the data to be encrypted leveraging the symmetry along the x axis and the asymmetry along the y axis that an elliptical function exhibits. ECC works on smaller keys, while providing the same level of security to data as RSA or Diffe-Hellman[18].

DNA encoding: DNA encoding based Cryptography can be defined as hiding data in terms of DNA Sequence. Just like the RSA and DES algorithms, in DNA Cryptology users use the DNA components with the plain text in a one-way function to generate the cipher text, which is similar to public key systems[7]. The components of a DNA structure are used as a lookup table corresponding to the symbols that are to be encrypted, to encrypt the plain text.

Quasi-group cipher (similar to Latin squares): this is a mathematical group that is used as a substitution box on the actual value acting as a lookup table[22]. It works similar to a lookup table but with multiple iterations.

Orthogonal matrix: in this approach, an orthogonal matrix of the same dimensions as the plain text is generated using a pseudo random number generator, and matrix multiplication is carried out with the actual frame data which is the plain text[23]. To decrypt the same, we multiply the cipher text with the inverse of the initial matrix.

3.3.2. Key generation techniques

Chaotic algorithms[28–32]

Chaotic systems/hyper chaotic systems: chaotic systems are based on Lorenz system is a system of ordinary differential equations[7,14]. The model is a system of three ordinary differential equations now known as the Lorenz equations. An extension of this is a hyper-chaotic system (5D, 12D)[19,24–26]. Hyper chaos is a higher dimensional chaotic system having two or more positive Lyapunov exponent. The number and the sign of the Lyapunov exponent of a dynamical system measure the rate of separation of infinitesimally close trajectories. Other systems that have been utilized and evaluated, Hitz-zele map and TinkerBell Map[27], with their own set of defining differential equations and initial values. The chaotic systems are generally a subroutine in the encryption technique and are used as pseudo random number generators for key generation for the encryption process[33].

Ikeda differential equation: time delay is inherent in physical systems. Time delay systems can be modelled by DDE that exhibit chaotic behavior. The dimension of the chaotic attractor can be increased with the value of the delay time. This feature takes its usefulness in generating random numbers and secure key generation in cryptography. The Ikeda DDE models a passive optical bi-stable resonator system[24,32,34,35].

Cellular automata

A Cellular Automaton is a discrete model consisting of a regular grid, of any dimension, with each cell of the grid having a finite number of states and a neighborhood definition[28]. There are rules that determine how these cells interact and transition into the next generation (state). The rules are mostly mathematical/programmable functions that depend on the current state of the cell and its neighborhood. Cellular automata are used here as a pseudo random number generator for key generation for the encryption process.

In summary, depending upon specific requirements, there is an array of encryption algorithms available that can be employed to achieve the desirable results. Different encryption algorithm offers different levels of security and based on how sensitive the video data that needs to be encrypted is, one can choose a relevant technique. Some of the techniques mentioned above offer lower time and space complexity compared to others which is also something to be considered when choosing an appropriate encryption algorithm for one's requirements. The different codecs available and different encoding standard in place regarding video data also need to be taken into consideration when selecting a technique.

4. Evaluation parameters of video encryption algorithms (RQ2)

This section is divided into three sub sections. The first section covers the different categories of evaluation metrics that are used to determine how well a video encryption algorithm performs. The section also gives a brief explanation of those categories. The second section covers the parameters that come under the above mentioned categories and it is divided into two sub sections. The first subsection talks about the evaluation metrics that are used frequently to assess the performance of the video encryption algorithm and the second subsection talks about parameters that are not used as frequently. The two subsections also provide a concise definition of the parameters to help understand their purpose in performance evaluation of encryption algorithm. The third section covers the different attacks that an encrypted video is susceptible to and the various techniques that are employed to accomplish the attack. It also gives a brief description of the different attacks.

4.1. Categories of performance evaluation parameters

The different evaluation parameters can be broadly divided into three categories. These three categories were chosen because they encompass the various factors that need to be taken into account when evaluating the performance of an encryption algorithm. **Figure 5** gives an overview of the different parameters and their respective categories. The categories are as following:

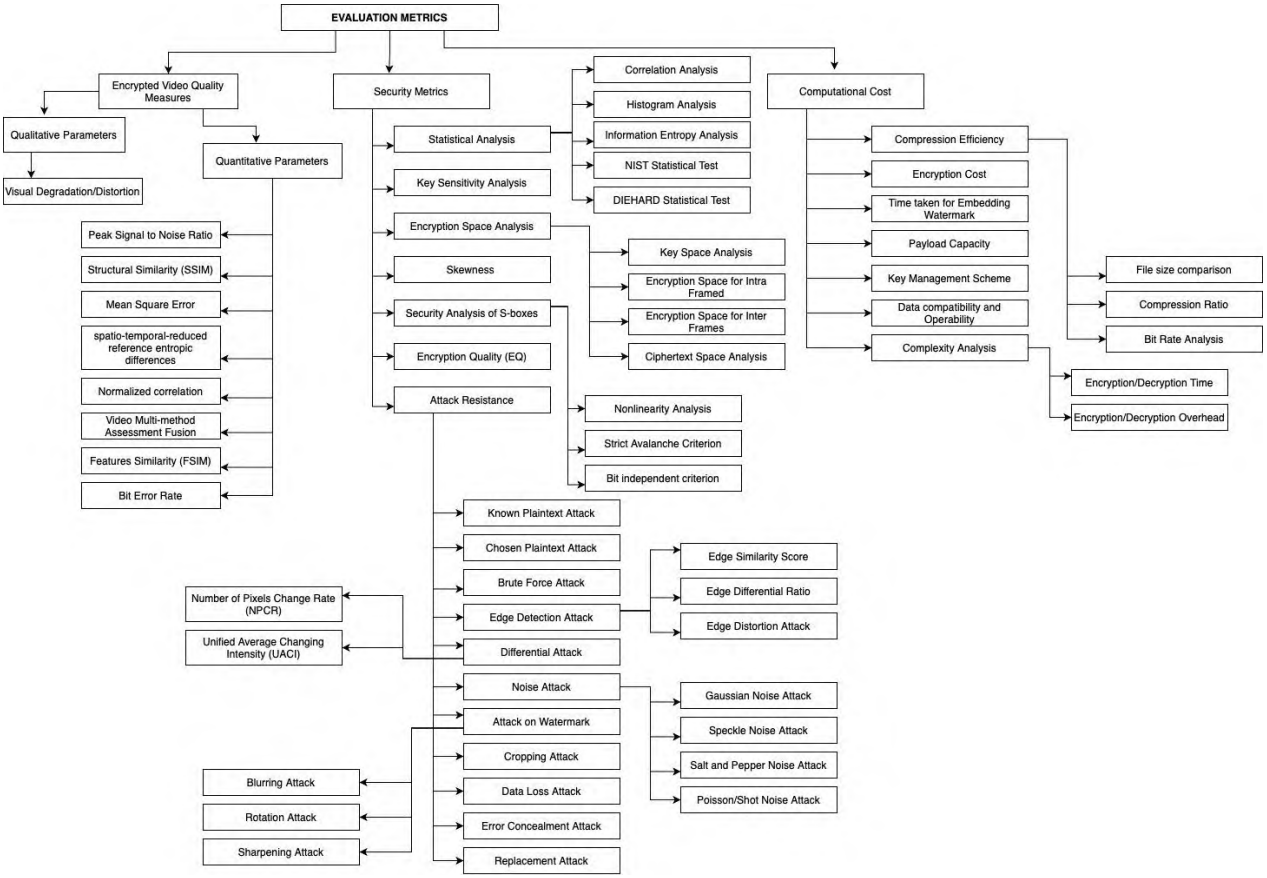

Figure 5. Categories of performance evaluation parameters.

4.1.1. Encrypted video quality measures

These metrics are used to assess the quality of the proposed encryption method. One of the basic requirements of an efficient video encryption algorithm is to make sure that the encrypted video is significantly different from the original video. These measures help quantify components of a video frame such as the structure, luminance and contrast and relationship between individual pixels. The value of these parameters helps decide how well an encryption algorithm performed in terms of how similar the encrypted video frame

is to its corresponding original video frame. The parameters in this category are also applied to the decrypted video in order to evaluate the quality of the reconstructed video and determine the similarity between the reconstructed video and the original video.

4.1.2. Security metrics

These metrics are used to determine how secure an encryption algorithm is. They allow the experimenter to evaluate how prone the algorithm is to different kinds of security attacks. Security metrics involve measuring the encryption space of the encryption method used in order to evaluate susceptibility to brute force attacks and also help quantify the randomness of pixels in the encrypted video by performing statistical tests on the encrypted video frames.

4.1.3. Computational cost

A basic encryption algorithm involves the generation of a secret key and the usage of that key to encrypt the video frames. However, depending upon the algorithm, the encryption method could involve the usage of chaotic maps for key generation or application of image processing techniques before the encryption process, which means that the cost of executing these algorithms would differ. These metrics help assess the amount of resources required to execute the proposed encryption method. Computational cost involves measuring the overhead generated while encrypting a video frame as well as the time and space complexity of implementing the encryption process.

4.2. Parameters

The following subsections cover the various parameters that are used to evaluate the different aspects of performance of an encryption algorithm. The frequently used parameter are the ones which are used to evaluate performance of the encryption method in three or more different papers. The rarely used parameters are the ones which are mentioned in less than three papers. **Table 1** shows the ideal values of the parameters discussed below.

Table 1. Various parameters and their ideal values.

Parameter	Ideal value
PSNR (Peak Signal to Noise Ratio)	Between 30 dB to 50 dB
SSIM (Structural Similarity Index Measure)	Between original and encrypted video frames: 0 Between original and reconstructed video frames: 1
Correlation coefficient	0
Information entropy	8
NIST statistical test	*P*-value should be in the range [0, 1)
FSIM (Features Similarity Index Measure)	Between original and encrypted video frames: 0 Between original and reconstructed video frames: 1
Skewness	−0.5 to +0.5 for a symmetric distribution.
DIEHARD statistical test	*P*-value should be in the range [0, 1)
VMAF (Video Multi method Assessment Fusion)	100

4.2.1. Frequently used parameters

Peak signal to noise ratio: the term peak signal to noise ratio (PSNR) refers to the ratio of the maximum possible value of a signal to the power of a distorting noise that affects the quality of its representation. The comparison of the PSNR values (expressed in decibels) of the original video frames and the reconstructed video frame after decryption help in quantifying the quality of the encryption method. In general, a higher PSNR value correlates to a higher quality image. PSNR can also be used to evaluate the quality of the encrypted video stream. As good encryption method would significantly result in decreased objective quality of the

encrypted video. A low PSNR value of an encrypted video frame correlates with the efficacy of the encryption method used PSNR. The mathematical Equation (1) to calculate PSNR is given below where MAX is the maximum value for pixels in a video frame and MSE is the mean squared error.

$$20log_{10}\frac{MAX}{\sqrt{MSE}} \tag{1}$$

The metric was recorded for different videos of different classes and at five different QP. The PSNR drops at QP 17 from 44.77 dB in average to around 10.17 dB. Similar PSNR values of encrypted videos are reached on different QPs values which implies that significant distortion was achieved as a result of the proposed encryption algorithm[4] used the PSNR metric to compare original and encrypted video frames as well the original and the reconstructed video frames. PSNR was recorded for six different videos with the PSNR ratios for encrypted videos close to the optimal value. The PSNR ratios for the reconstructed video were found to be infinite for all six videos which indicates that the video was reconstructed perfectly after decryption.

Structural similarity index measure: SSIM is used for measuring similarity in the structure of two images. It is a metric that quantifies changes that occur in the structural information of a video frame after processing such as compression or encryption. The pixels in an image or a video frame have strong inter-dependencies, especially when they are spatially close. These dependencies carry information that could allow us to know about the structure of the objects in the visual scene. Ideally, an encrypted video frame would bear no similarity to the original video frame and a reconstructed frame would be exactly the same in structure as the original frame. The SSIM value ranges from 0 to 1, where 1 means a perfect match of the reconstructed video frame to the original one. SSIM between two windows x and y of common size N × N is calculated using the below mentioned Equation (2) where μ_x is the pixel sample mean of x, μ_y is the pixel sample mean of y, σ_x^2 and σ_y^2 are the variances of x and y respectively and c_1 and c_2 are two variable to stabilise the division with weak denominator.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{2}$$

El-Shafai et al.[4] used SSIM to measure similarities between the original and encrypted video frame as well as the original and the decrypted video frames. The metric was recorded for six different videos. For the encrypted video frame, SSIM values span across 0.0019 to 0.0059 which is close to the recommended value for sufficient distortion. The SSIM findings between the original and the decrypted videos were 1 for all six videos which indicates that the video was perfectly reconstructed.

Correlation analysis: correlation analysis is performed on the encrypted video frames to measure the relationship between two variables in a video frame. It is a form of statistical analysis that assesses how adjoining pixels in an image correlate to each other. Digital video frames usually contain a large amount of redundant information which results in a strong correlation between the pixels of a video frame. This correlation of adjacent pixels is expected to be reduced after the application of an encryption algorithm. Correlation is measured in three different directions in a video frame: Horizontal, Vertical and Diagonal. Complete correlation between two variables is expressed by either +1 or −1 and complete absence of correlation is represented by 0. An ideal encryption method would result in a correlation value of 0 for an encrypted video frame. If *x* and *y* are two variables, then the correlation $\rho(x,y)$ between them could be calculated using the below Equation (3) where *cov* is the covariance between *x* and *y* and σ_x and σ_y are mean of *x* and *y* respectively.

$$\rho(x,y) = \frac{cov(x,y)}{\sigma_x\sigma_y} \tag{3}$$

Kordov and Dimitrov[27] use an encryption model with two chaotic systems for pseudo random generation of cipher keys. Five videos were tested with the proposed scheme and correlation coefficient was calculated

for all five videos. All encrypted files had values very close to 0, which means the adjacent pixels' values have no dependence, indicating strong encryption.

Histogram analysis: a histogram analysis can be done on the grayscale value or the RGB value of the pixels in a video frame. The histogram is a graphical representation that shows the pixel intensity value. For any given encryption technique to be efficient, the histograms of the original input video frame should differ completely from the encrypted video frame while the histograms of the original input image and the reconstructed image should be the same. Generally, the pixel distribution of an original video frame is random whereas the same for an encrypted video frame is uniform.

Information entropy analysis: entropy refers to the measurement of uncertainty in information theory. In the context of a video, entropy is used to determine the probability of the appearance of certain pixels in a given video frame. This metric is used to assess the security aspects of encryption algorithms that make use of chaotic maps and pseudo random number generators for the generation of an encryption key. The efficiency of chaotic encryption is based on the ability of some dynamic systems to produce sequences of numbers that are random in nature. This sequence is then used to encrypt a message. Information entropy is a way to quantify the degree of randomness of the sequence used as a key. Greater the quality of random number generation results in a greater quality of the random keys produced which furthermore results in higher security of the secret key. For a truly chaotic system, the value of information entropy is 8. Given a discrete random variable *X*, distributed according to p: $X \rightarrow [0, 1]$, the information entropy *H(X)* is calculated using the below Equation (4).

$$H(X) := -\sum_{x \epsilon X} p(x) \log p(x) = E[-\log p(X)] \quad (4)$$

Kordov and Dimitrov[27] recorded information entropy of the first and the last frame of five different videos and all values are around 7.99 which is very close to the perfect value of 8.

Key space analysis: the key space of an encryption technique is the set of possible encryption keys that can be used to encrypt the video data. Key space analysis is conducted on an encryption method to check the susceptibility of the method to the brute force attack. Cryptosystems have a natural limit to the number of keys used to encrypt due to the nature of the rules in place and an algorithm with a larger key space is considered to be more robust in context of a brute force attack. A key space is calculated by utilizing n bits to create the key[27]. The encryption process employed chaotic maps and PRNG, resulting in a substantial key space capable of withstanding brute force attacks. Additionally, Waseem and Khan[29] implemented a quantum video encryption algorithm that incorporated qubit-planes controlled-XOR operations and enhanced the logistic map for improved security. In the proposed method, the total number of keys is determined by the control parameters and the initial values of three independent improved logistic maps employed. The key space for each individual key is calculated as 32,768, resulting in a substantial overall key space of approximately 1.23794003928538 x. This extensive key space provides sufficient protection against exhaustive search and other brute-force attacks.

Key sensitivity: an encryption is typically a string of bits generated to scramble data so that said data is not easily comprehensible. The basis of an efficient key generation algorithm is to produce a key that is unique and unpredictable. Since the efficacy of the algorithm is highly dependent on the encryption remaining a secret, an ideal encryption algorithm would be highly sensitive to the secret key. This means that even the slightest change to the original key should produce a completely different encrypted video frame. The sensitivity of the encryption key to the initial conditions can be measured by analysing the results of encryption by encrypting the original video frame with the two keys that are just one bit different. Histogram Analysis/PSNR/SSIM used to quantify.

NIST statistical test: the statistical test suite developed by NIST (National Institute of Standards and Technology) is an exhaustive document that outline the various aspects of randomness in a long sequence of

bits. The NIST has documented 15 statistical tests and a sequence is verified against each test. At a test for one sequence, a statistics called *P*-value is calculated from the tested sequence. *P*-value represents the probability that an ideal random number generator would produce less random sequences than the sequence being tested. For every test to be successfully passed, the obtained *P*-value needs to be in the range [0, 1). Kordov and Dimitrov[27] and Haridas et al.[22] conducted NIST tests on their respective sequence and both methods qualified the tests.

Complexity analysis: for an encryption algorithm, complexity is measured by two different metrics. Encryption Time refers to the time required to process the original video data under the proposed encryption algorithm. Similarly, decryption time refers to the time required to reconstruct the original video data. Another way to measure complexity of an encryption algorithm is by measuring the overhead. Overhead refers to additional computational power that is required to process and transmit encrypted data as well as the power required to decrypt the data once it is transmitted. Gautier et al.[21] used a selective encryption algorithm and calculated the overhead only for the decoder since the encryption overhead was found to be negligible with respect to the encoding time. Decryption time and decryption overhead were recorded for different classes of video at three different QPs. The decryption time does not exceed 3 seconds even for high bitrate and high resolution videos and the average deciphering overhead remains lower than 4.23%. El-Shafai et al.[4] used an optical bit-plane based cryptography algorithm and recorded the enciphering/deciphering time for six different HEVC frames. The shortest time for a frame was found to be 5.5 seconds and the longest time for encryption was found to be 7.9 seconds. Waseem and Khan[29] used a quantum video encryption algorithm and the complexity was calculated based on the number of the basic quantum gates used in the entire encryption process.

Compression ratio: compression ratio is defined as the ratio of the number of bits in an uncompressed video to the number of bits in a compressed video. Kelur et al.[10] used Harr–cascade classification technique to selectively encrypt the objects. The encryption of selected object is achieved by RNS (Residue Number System) operation and achieved a significant compression ratio due to the use of RNS which stores only residues of respective pixels. Karmakar et al.[7] used a sparse representation based compressive video encryption using hyper-chaos and DNA coding and was able to achieve the goal to obtain higher compression of video-frames without degrading the reconstruction quality.

4.2.2. Rarely used parameters

Features similarity index measure: FSIM is a video quality assessment metric that maps out the features such as light variation, phase congruency and gradient magnitude, of a video frame and measures the similarity in those features. FSIM was proposed as a quality assessment metric based on the fact that the human eye perceives a visual mainly according to its low-level features. El-Shafai et al.[4] used the HEVC codec and used FSIM for analyzing the encryption/ decryption competence of the proposed optical HEVC cryptography algorithm. It determines the value of local similarity among two distinct HEVC frames. In the security analysis, it is tested amongst the enciphered and original HEVC frames, and amongst the deciphered and original HEVC streams. For a good encryption algorithm, the FSIM value among the original and the encrypted frame is lower and the same is higher among the original and the deciphered frame. The quantity of the FSIM metric is in the decimal range achieve an average FSIM of 0.3 between the original and the enciphered frame and a value of 1 for all six frames for the similarity between original and the deciphered frame.

Skewness: skewness is a statistical metric used to measure the degree of asymmetry observed in a probability distribution. The distribution of an original video frame is highly varied. However, the distribution should be symmetric for an encrypted video frame so that an attacker is not able to guess any parts of the frame based on its distribution. The skewness should vary between −0.5 to +0.5 for the symmetric distribution. Skewness of a distribution is calculated using the below Equations (5):

$$Skewness = \frac{3(Mean - Median)}{Standard\ Deviation} \quad (5)$$

Karmakar et al.[7] used hyper chaos and DNA coding for video encryption and employs skewness as way of assessing the security of the algorithm. The encryption method was applied on a single original frame of three different videos and the skewness varied between −0.2 to +0.2. The variation in the skewness of the original video frame was large. After encryption, the variation in the skewness ranges between 0.2 to −0.3 and thus qualifies the security test.

Video multi-method assessment fusion: VMAF is a perceptual video quality metric that predicts the perceived quality score of a video sequence. VMAF values span from 0 to 100 where a score of 100 indicates a good perceptual video quality and 0 refers to a very low perceived video quality. Gautier et al.[21] employed a selective encryption algorithm and used VMAF to assess the quality of the encrypted video. The proposed algorithm resulted in large degradation of the subjective video quality. The VMAF score was recorded for encrypted frames of different video classes at five different QP with the lowest average score of 9.32 observed at QP 17 and the highest average score of 10.66 observed at QP 32.

DIEHARD statistical test: similar to the NIST Statistical Tests, the DIEHARD tests are a set of tests designed to measure the randomness, hence the breakability and the usability quality of a random number generator. DIEHARD software performs 19 tests for randomness evaluation of the produced binary sequence and for every test to be successfully passed the obtained *P*-value needs to be in the range [0, 1). Kordov and Dimitrov[27] used video encryption model based on two different chaotic systems and pseudo random generation of keys. The DIEHARD Statistical test was used to assess the quality of the proposed PRNG and the binary sequence generated passed all 19 tests.

4.3. Different attacks on video encryption algorithm

A cryptographic attack is the method of circumventing the security of the encryption algorithm by exploiting the weakness in the cipher key, the cryptographic process itself or the key management scheme. There are various kinds of attacks that can be employed on an encrypted video frame in order to extract the original frame. This section aims to provide an explanation of these attacks.

4.3.1. Known-plaintext attack

The goal of the known-plaintext attack is to derive the cipher key used to encrypt the original message. It relies on the assumption that the attacker has access to the ciphertext and its corresponding plaintext. The attacker tried to guess the cipher key from the matching ciphertext-plaintext pair. Recovering the key allows the attacker to decrypt other ciphertexts that were encrypted using the same key. Encryption algorithms that are based on simple ciphers such as the substitution cipher and the simple XOR cipher are the most susceptible to known-plaintext attacks. Valli and Ganesan[24] used two different schemes using 12D maps and Ikeda time delay system for video encryption and checked resistance against known plaintext attack by introducing two kinds of external disturbances in both schemes. The first is a frame drop technique in which one frame is dropped and the subsequent frames are decrypted. The second is a frame swap technique in which swapping of frames is induced in the encrypted sequence of frames and the new sequence is then decrypted. In case of Ikeda DDE, there was unintelligible decryption and hence the scheme is robust against known plaintext attack. Li et al.[17] used a video encryption scheme based on the Cloud-Fog-Local architecture and claimed robustness against known plaintext attacks by stating that the proposed algorithm has a large enough ciphertext space. A large amount of data makes it difficult for attackers to guess the whole plaintext.

4.3.2. Chosen-plaintext attack

While executing the chosen-plaintext attack on a cryptographic system, the attacker can choose arbitrary plaintext data to be encrypted and receives the corresponding cipher. The attacker, then, tries to acquire the

cipher key or alternatively, tiers to develop an algorithm that allows him to decrypt other plaintexts encrypted using the same cipher. The chosen-plaintext attacks are more efficient than known-plaintext attack because it allows the attacker to obtain more information about the cipher key and the whole attacked system based on any kind of input data. Any cipher that is vulnerable to the known-plaintext attack is automatically vulnerable to the chosen plaintext attack as well. Xu et al.[19] used a video encryption method based on cross-coupled chaotic cipher and checked the resistance of this scheme against the chosen plaintext attack. The average hamming distance is used to calculate the difference between the two key streams used to encrypt different video files under the same initial key. In the most ideal case, the value should be equal to 0.5. The method was able to achieve the value of 0.4989 which is very close to the ideal value. Valli and Ganesan[24] used a chaos based video encryption using maps and Ikeda time delay system. The proposed scheme is tested with the two plain video frames P and Q, encrypted with the same key and the corresponding cipher frames are denoted as P1 and Q1. A mask frame M is generated by performing XOR operation on P and P1. With the help of this mask frame, Valli and Ganesan[24] tried to retrieve Q by using the XOR operation on M and Q1. This is found to be unsuccessful using the 12D chaotic map and the Ikeda DDE

4.3.3. Brute force attack

The brute force attack is an attack model that uses trial and error in an attempt to guess the cipher key used in the encryption algorithm. The attacker methodically checks all possible cipher keys in the hopes of eventually finding the correct one. An encryption scheme that does not have a large enough key space i.e., the number of possible cipher keys that can be used to initialize the encryption algorithm, is the most susceptible to brute force attack. Any encryption algorithm with at least 128 bits as encryption key is considered as resilient to brute force attack. Xu et al.[19] used a cross-coupled chaotic cipher for encryption. The initial key k′ used to calculate the encryption and decryption synchronization vector consists of 512 bits and the two keys x′ and y′ of the chaotic system. If stored in a double-precision data type, each of them occupies 64 bits. Therefore, the key space of the encryption algorithm was much larger than and hence the scheme is robust against a brute force attack. Waseem and Khan[29] used a quantum video encryption algorithm based on qubit-planes controlled-XOR operations and improved logistic map and the overall key space of the proposed encryption algorithm is about =1.23794003928538 x which is large enough to withstand an exhaustive search.

4.3.4. Edge detection attack

In an edge detection attack, the attacker tries to obtain information about the original video frame, specifically edge information by analysing the edges of the corresponding ciphered video frame. There are several ways in which you can quantify susceptibility to an edge detection attack. Edge Differential Ratio is used to enumerate the difference between the edges of the original video and the edges of the encrypted video frame. A value of 1 or close to 1 implies that the edge information in the ciphertext is small. Similarly, Edge Similarity Score is used to measure the similarity between the edges of the original and the encrypted video frame. Edge Distortion is a way of measuring alteration of the edges of the original video frame as compared to its corresponding encrypted frame. Gautier et al.[21] used a selective encryption scheme and recorded an EDR value of 0.87 which shows the ability of the proposed method to hide edges and structural information in the encrypted frames. Xu et al.[19] used an encryption method based on cross-coupled chaotic cipher and was able to achieve an EDR value of 0.96, which implies that the edge information in the cipher text is very small.

4.3.5. Differential attack

Differential attack is similar to known-plaintext attack in the sense that the attacker has access to the original video frame as well as the encrypted one. The attacker tries to change the pixels of the original video frame and observes the changes that causes to the encrypted video frame. This allows the attacker to obtain information about the cipher used. There are two metrics that can be used to quantify susceptibility to differential attack. Number of Pixels Change Rate (NPCR) is defined as the change rate of the number of pixels

of the ciphered video frame when only one pixel is changed in the original video frame. Unified Average Changing Intensity (UACI) is used to measure the average intensity of differences between the original and the ciphered video frame. A high Non-Linear Pixel Change Rate (NPCR) and Uniform Average Changing Intensity (UACI) indicate that the encryption method utilized is highly resistant to differential attacks. Ideally, the NPCR value must be greater than 99% and the UACI value should be around 33%.

4.3.6. Noise attack

Noise attack analysis is performed in order to check the resistance of the enciphered video stream to the different kinds of noise. The enciphered stream is attacked with noise and the deciphered stream is then observed. This helps to determine whether or not an encrypted video will be properly reconstructed to its original form in case it gets intercepted with noise during transmission. There are different kinds of noise that can be applied to the encrypted image. The most common is the Gaussian Noise Attack. To observe the effects of gaussian noise attack, a normally distributed random value is added to each pixel and the resulting deciphered and enciphered HEVC frames of the Gaussian noise is studied with diverse variance rates. PSNR and SSIM is measured for deciphered frames to assess the quality of reconstruction. Speckle Noise Analysis is done to gauge the effects environmental conditions can have on the video sensor during video acquisitions. Similarly, Salt and Pepper noise analysis is done to measure the effect sudden and sharp disturbances can have on a video signal. It presents itself as sparsely occurring white and black pixels. Another form of noise analysis that is done on a deciphered stream is the Poisson/Shot Noise analysis. Poisson noise is a form of uncertainty associated with the measurement of light. It is studied as a variation in the sensed photon number at a particular exposure degree.

4.3.7. Attack on watermark

One of the techniques used when encrypting a video is watermarking, which is the process of hiding data in video frames and can be used with cryptography mechanisms to provide more security. Analysing attacks on watermarked frames allows the experimenter to check resistance of the proposed method against quality loss while deciphering a watermarked frame. The Sharpening Attack is applied to a watermarked frame to highlight the details of the image in the selected frame. It can enhance the changes in high and low frequencies in a frame. Another form of attack that is applied on a watermarked frame is the Rotation Attack. Rotation Attack is the process of rotating a video frame by a degree which distorts the watermark as the original position is changed. Similarly, in a Blurring Attack, a random sequence of real values is added to all frames of the watermarked video and is used to check the motion of the watermarked frame. Sharma et al.[6] used a frame selection mechanism followed by watermark embedding and the robustness of the proposed technique is tested against the various noise attacks mentioned above. The robustness of the technique entirely depends upon the values of PSNR, SSIM, NC, and BER. The proposed method was found to be robust against Gaussian Noise Attack, Sharpening Attack and Blurring Attack, since the average values of PSNR, NC, and SSIM decreased with an increase in attack value while BER increased with increase in attack value. However, the technique did not achieve good results against Rotation Attack.

4.3.8. Cropping attack

The cropping attack study is used to evaluate the potential of restoring and decoding original HEVC frames from enciphered HEVC frames in the event that a certain percentage of the encrypted frames has been vanished or obstructed. Cropping attack helps in assessing the robustness of an encryption method against faulty HEVC communication. It is similar to a data loss attack wherein information is stolen from a system without the knowledge of or the authorization of the system's owner. The effect of cropping or data loss attack can be observed by occluding a certain percentage of the encrypted stream and attempting to reconstruct the video from the remaining percentage. El-Shafai et al.[4] tested for robustness against a cropping attack and observed that the HEVC streams can be decrypted in a comprehensible form even if some parts of enciphered

HEVC streams are cropped in distinct and separate localities throughout the HEVC communication. This confirms resistance to cropping attack.

4.3.9. Error concealment attack

Error concealment attacks are a common form of attack for all selective encryption methods. Since selective encryption involves encryption of a subset of the video stream, it is always a possibility that some information might leak from the unencrypted bits of the video. An attacker can try to carry out error concealment attack by replacing all the encrypted bits with a fixed value and try to reconstruct the original video from the unencrypted bits. One of the scenarios to test robustness against error concealment attacks is to replace all the encrypted bits with zero and then decipher the encrypted stream. A low PSNR and SSIM value of the deciphered stream would suggest that the encryption method used is not susceptible to this kind of attack. This kind of attack is also referred to as replacement attack. Gautier et al.[21] checked for robustness against error concealment attack by replacing all encrypted bits with zero and decrypting the resulting stream. PSNR, SSIM and VMAF were recorded for this decrypted stream at five different QP and the obtained quality scores are low as expected. This confirms that the proposed method is robust against attacks based on replacement bits.

5. Challenges and limitations of video encryption algorithms (RQ3)

This section discusses the challenges and limitations of existing video encryption methods. Some of the solutions to these problems are briefly outlined as well. In the end, a summary and analysis of newfound insights into the connections between various forms of attacks and important aspects of performance are presented.

Dealing with time complexity and spatial complexity was one of the main problems with prior video encryption techniques from year 2017 and 2018. In the past, several methods just encrypted one frame at a time without using video compression, which produced an excessively high temporal complexity. Due to a lack of focus in this area, dealing with numerous attacks, including geometric attacks like rotation, and flipping as well as plain text attacks, was another significant issue. Additionally, it appeared that the algorithms had little effect on crucial performance factors like pixel change rate and value. Thus, with video encryption techniques, there was always a trade-off between speed and security. Most of the articles restricted their application to non-cloud applications.

The earlier challenges like expensive short-pulse laser (operation only in active mode) and dependence on the precise repetition of the ultrafast event during the captures (multi-shot imaging), inability to image luminescent transient events, monochrome scaled captures, low number of captured frames (short duration of recording), demanding storage and transmission capacity requirements, extremely high built costs, high maintenance, and oversized dimensions is now fulfilled by the CCRM camera[30]. Amplitude encoding has been added to the encryption and compression stages, considerably expanding the key-space and reducing the risk of brute force attacks on data recovery.

One of the striking methods was the which is not only limited to grayscale images but can also be extended to colour images in a short amount of time making it suitable for today's fast communication[25]. However, hardware still remained an issue.

Video selective encryption has also advanced over time, to the point that it works with any video codec by encrypting only the data immediately preceding the start code, which was formerly part of the video bitstream[14]. Further, unlike the pre-existing SEAs that are codec-specific, the video codec-independent encryption framework allowed for compliance with the standard video codec. Secondly, the proposed solutions prevented users from seeing any undecipherable video content. Existing SEAs, on the other hand, only allow for limited data gathering from distorted decoded video. As a bonus, the proposed approaches based on the

start code are more useful and practical since they prevent decoding, which improves security. Over the years and various approaches, people have tried to increase visual security[31]. Improved visual scrambling effect, less video file size, and resistance to the MBS sketch attack and standard JPEG sketch attacks are only a few of the benefits of a new selective encryption strategy for H.264/AVC.

The problem of 4K-UHD problem- HEVC 4K videos and backwards compatibility with lower resolution video is addressed in where sparse selective encryption for HEVC 4K video is done using spatial error spread[20].

Early methods of video encryption did not involve the use of compression. Later, as described in, we were able to accomplish our goals of obtaining better compression of video-frames without reducing the quality of their reconstruction and inventing an effective encryption technique concurrently[7].

Kelur et al.[10] and Alattar et al.[30] address the problem of time. Kelur et al.[10] propose a method where the size of the original image is reduced while Alattar et al.[30] come up with a solution by using I macroblock as encrypting every other I-macroblock requires about half the processing time of encrypting all I-macroblocks. Li et al.[28] resolve the inherent problem of 2D CA mask-based encoding methods.

The cloud based techniques involves Separable reversible data hiding and encryption for HEVC video while the real time implementations involve algorithms like HEVC Selective Encryption of CABAC, with H.264/AVC and other video encryption standards with high efficiency, Chaos based video encryption using maps and Ikeda time delay system, A new approach to digital content privacy using quantum spin and finite-state machine.

Table 2 shown below lists the major attacks observed and the parametric requirements to prevent these attacks.

Table 2. Attacks on video encryption algorithms and its mitigation.

Attack	Mitigation
Brute force	Large key space should be used, high key sensitivity
Entropy attack	IE near to theoretical value of 8
Statistical attack and visual analysis attack	Histogram should be uniformly distributed
Differential attack	NPCR, UACI, Hamming distance-all close to optimal value
Known plain text attack	Low PSNR and SSIM
Chosen plain text attack	Hamming distance = 0.5
Edge detection attack	Gaussian–Laplacian, edge difference ratio close to 1
MBS sketch attack	Number of bits of different macroblocks tend to be nearly equal in an encrypted video, or the positions of most macroblocks are randomly permuted
Sketch attacks	Edge similarity scores (ESS) should be low

The papers cover a myriad of topics related to encryption of videos, these include approaches towards encryption of video via encryption of individual frames, encryption of encoded bits that are outputted by different codecs, encryption before encoding, encryption after encoding, encryption using hardware techniques etc. The papers also discuss various key generation techniques such as pseudo random generation via different methods such as chaotic systems, cellular automata, and usage of different encryption algorithms for encryption such as AES, DES, Rabbit, RSA, Elliptical Curve Cryptography etc.

The methodologies implemented differ in a number of ways, such as the sequence in which the individual techniques have been employed, the parameters that have been chosen to be encrypted, the efficiency of the technique that has been chosen for parameter or frame selection, the system of chaotic equations used for encryption and its complexity, types of transforms used etc. These factors further add to the time and space complexity of performing the encryption operation as well as the level of security that is provided by the methodology and the susceptibility to different types of attacks.

All of the approaches that we have covered are aimed towards confusion and diffusion of the initial video data, be it via the usage of transforms and encryption algorithms. Confusion is employed as it hides the relationship between the encrypted data and the key whereas diffusion is employed for increasing the redundancy of the data and hide the statistical relationship between the ciphertext and the plain text.

6. Discussion and conclusion

With the advent of technology from portable recording devices, development of cameras in our mobile devices to the leaps in the data compression, encoding, telecommunication and the real time connectedness that the internet brings us has led to an explosion of video data that is being shared and consumed all over the world. Be it movies being viewed on a streaming platform, video calls, live streams of different categories, industrial training videos being shared and viewed, videos being posted on social media, video has fast risen to be the preferred type of media to be consumed. With this explosion in the amount of video data, security and confidentiality of multimedia contents has gained prominent importance in the majority of applications to ensure safe storage and transmission of this data. This is where the concept of video encryption comes in. video encryption is a process of digitally hiding videos to ensure that unwanted interception and access to the transmitted videos is prevented. The process involves encrypting videos using encoding software and hardware to secure the data.

As video data is quite massive (video being a huge collection of frames/images), it is not practical to encrypt every frame and data point in the video since it will be not be cost effective in terms of both time and computing resources. To accommodate this limitation while ensuring adequate safety, different types of video encryption methodologies come into play that allow encryption of specifically selected frames or data points from the video, providing the highest level of distortion of the video, without encrypting the entire video data.

Applications of video encryption are quite all encompassing as virtually all kinds of applications and devices use video data. The safe transmission of the same while ensuring low time and space complexity is of utmost importance. An example of this is a streaming service. The video needs to be encrypted in a format that is safe from attackers while being deciphered in real time to decrease latency and load times and ensure the highly availability promised by video sharing and streaming services. In some cases where high level of security is the main requirement and longer load times are not as much of an inconvenience, much more complex video encryption methodologies can be utilised, which would ensure high level of security. Use of this kind of methodologies might include encryption of CCTV footage, court recordings, industrial training videos etc.

The motivation for this paper was to create a classification framework of the existing video encryption methodologies and video encryption validation methodologies, that can be used by any practitioner as a reference to the vast field of video encryption and on the basis of their requirement, be able to pinpoint the type of encryption that would satisfy the requirement of their use case.

The different video encryption techniques can be broken down into two major functions, the first includes transforming the video data into its most important components via mathematical transforms, hardware-based techniques or encoding of the video data through standard encoding techniques. The second is the encryption of the transformed data (which is easier to work with than the raw video data), by using an encryption function

such as the XOR with the transformed data and the encryption key (generated by a key generation technique such as chaotic system, cellular automata etc.) as input. Measuring the performance of these different techniques is an important part since it allows the practitioner to effectively evaluate the method used and whether or not they were able to achieve optimal results. Evaluation of these techniques can be done on the basis three main aspects. First is the quality of an encrypted video. Sufficient distortion is a primary requirement of an encryption algorithm and encrypted video quality measures allows for a way to quantify the distortion in video data after encryption. Second is the level of security the encryption algorithm provides. The security parameters allow practitioners to measure the susceptibility of their chosen algorithm to unwanted interceptions. Third is the computational cost of a video encryption algorithm. These measures give a way to evaluate the time and resources needed to apply a chosen encryption algorithm. The originality of the paper stems from the fact that this paper does not aim to list the most relevant or best performing video encryption techniques, but to act as a map for a user to understand what type of encryption they are looking for and what validations it needs to fulfil to achieve the level of encryption of video data that they require.

In summary, after evaluating the different methodologies used for the process of video encryption, what observed is that depending on the end goal of the practitioner, there are different encryption techniques that one can make use of. In the case where the practitioner requires videos to be encrypted while being transmitted to a large audience (for example, Streaming Platforms), the practitioner will need to keep in mind the various codec and standards in place and choose the methodology accordingly. In this case it will be effective to utilise one of the approaches categorised under post encoding techniques, as these techniques do not require additional compression or encoding of data and will fit in the standard encoder decoder architecture used in video players, ensuring that the resulting data is compatible for the majority of the users and available for mass distribution without an issue. These techniques will also have lower time and space complexity, as the encryption process is already being carried out on data that has already been compressed and encoded by the encoding module as per the video encoding standards defined by MPEG, so there will not be a requirement for performing additional transforms on the data to optimise the encryption process. For pre-encoding methodology, on the HEVC or H.264 encoding standard (these comprise of 97% of the video traffic), the NAL unit which has the highest effect on encryption are the motion vectors, as they generally have the highest spatial influence. So, it is advisable to make sure that this unit is encrypted in addition to any other NAL unit, for the highest distortion from the original data in the encrypted data. If the requirement of the practitioner is high security of video data that will be accessed a limited number of times by a small audience (For example, industrial applications, CCTV cameras, traffic camera, clandestine observation data, court or government proceedings etc.), pre-encoding methodologies would be useful. These methodologies will take the raw video stream, break it down into frames, and apply mathematical transforms on them, either on each frame or on specific frames depending upon frame selection technique that is employed by the practitioner. The mathematical transform will generally have a twofold application, i.e., it will compress the data and augment the distortion of data on top of the encryption technique that will be used. After this, the selected encryption algorithm and key generation technique will be used to completely encrypt the data and pass it on to the codec, for further transmission. As the pre-encoding process processing of the uncompressed, raw video data, which is generally quite large, this approach will have an increased time and space complexity in comparison to post encoding approach. Hence, this approach will not be advisable for use cases where requirement is high availability with low latency.

Apart from the different encryption techniques that can be used, another major aspect of video encryption is the evaluation of these different techniques. Since the different methods of encryption employ different methodologies to achieve encryption, there is a wide range of parameters that can be used to evaluate the performance of the algorithm. Encrypted video quality measures such as PSNR provide a highly accurate way to quantify the video quality after encryption. PSNR allows practitioners to measure the level of distortion in an encrypted video. Since one of the basic requirements of an effective encryption algorithm is significant

distortion, PSNR value proves to be a reliable indicator of efficiency. SSIM is another quality measurement parameter that gives practitioners a way to measure structural similarity between the pixels in a video frame. Since, it is expected for pixels to have low inter-dependencies, especially when they are spatially close, after encryption as opposed to before encryption, SSIM values are extremely helpful since they allow practitioners to measure the inter-dependencies between pixels. Another parameter that is useful when it comes to measuring the level of security provided by the encryption algorithm is correlation analysis. It is a form of statistical analysis that assesses the correlation between two adjoining pixels in an image. correlation of adjacent pixels is expected to be reduced after the application of an encryption algorithm. Correlation is measured in three different directions in a video frame: Horizontal, Vertical and Diagonal. Evaluating key sensitivity is also a reliable way of measuring security level of an algorithm. Since the efficacy of the algorithm is highly dependent on the encryption remaining a secret, an ideal encryption algorithm would be highly sensitive to the secret key. This means that even the slightest change to the original key should produce a completely different encrypted video frame. Other than video quality and security level, computational cost is also a significant aspect of evaluation of an encryption algorithm. Complexity analysis helps assess the amount of resources required to execute the proposed encryption method.

Encryption requires extra computation in addition to the encoding process (which is a necessity for efficient transmission), which makes it a time sensitive process and a difficult problem to overcome when it comes to resource constraint environment such as mobile computing. This can lead to a trade-off between security and computational efficiency[1]. It can also lead to loss of significant information or depreciated video quality on the receiver's end in case an issue arises during the decryption or decoding process. Encryption can be very error sensitive. For encryption methods wherein the video frames are encrypted before they are encoded, one needs to ensure that the transformations that are performed as part of the encryption methods results in a format that is compatible with the standard codecs. For encryption methods that are dependent on encryption of principal bits outputted by the codec, the encryption method will be highly specific to the codec itself and might not be applicable in case said codec is changed. Encryption of live stream data is one of the largest challenges due to its real time nature. The application of genetic algorithms to video encryption techniques in order to optimise the value of parameter such as UACI and NPCR or optimise the selection of video frames to be encrypted remains an unsolved problem today. Even though encryption processes have certain limitations as discussed above, there are certain advantages that cannot be overlooked. Encryption allows for secure transmission of highly sensitive and confidential information. Unwanted interception of this information can have major repercussions, ranging from threat to personal safety to threat to national safety, and protecting such information is of utmost importance. Encryption allows people to protect their copyrighted content and helps to prevent plagiarism. It provides protection to a person's intellectual property and prevents others from exploiting the copyrighted content without the copyright holder's permission. Encryption plays an important role when it comes to monetizing content by protecting the data from falling into unauthorized hands.

Our paper aims to provide a direction to practitioners in order to help them identify which kind of encryption techniques would be best suited to their specific requirements. The bifurcation of the techniques can help them evaluate their use case and select a process that meets their pass criteria for optimal encryption and the discussion on performance metrics lays down groundwork that can help them evaluate the efficiency of their methodology. The discussion on limitations will allow practitioners to better understand the scope of a particular encryption method and provide with them all the information needed to make an informed decision. The paper can be a one stop for identification of the type of algorithm or technique and will provide an introduction and a jumping off point for further research or implementation a practitioner is looking for.

Future works can focus on a new encryption methodology that is a combination of pre-encoding and post-encoding techniques to provide an even greater level of security to the video data. According to this method, the data will first be encrypted before ingestion at the encoder by one of the pre-encoding techniques, after

which the data will be encoded and the encoded data will then be encrypted by a post encoding methodology. This would lead to two rounds of encryption and provide a much higher level of security, though it might cause an uptick in the overall space and time complexity. Adding to the above, the encryption algorithm and the key generation techniques that are the most used are the AES and different versions of chaotic systems respectively. AES is a world standard for encryption algorithms and provides a very high level of security in comparison to other methods. Depending upon the user's requirement such as level of security, complexity, speed a practitioner can use other ciphers too. For example, if the requirement is high speed encryption with a comparatively lower level of security Rabbit is another cipher that is widely used, while if the requirement is cipher that requires less computational power, PRESENT can be used.

Author contributions

Conceptualization, AK, AS and PC; methodology, PC; writing—original draft preparation, AK, AS and SA; writing—review and editing, PC and SA; supervision, PC.

Conflict of interest

The authors declare no conflict of interest.

References

1. Rabieh K, Mercan S, Akkaya K, et al. Privacy-preserving and efficient sharing of drone videos in public safety scenarios using proxy re-encryption. In: 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI); 11–13 August 2020; Las Vegas, USA. pp. 45–52.
2. Al-Thahab OQJ, Hussein AA. Implementation of stego-watermarking technique by encryption image based on turbo code for copyright application. In: 2020 1st. Information Technology to Enhance e-learning and Other Application (IT-ELA); 12–13 July 2020; Baghdad, Iraq. pp. 148–153.
3. Abomhara M, Zakaria O, Khalifa OO, et al. Enhancing selective encryption for H.264/AVC using advanced encryption standard. *International Journal of Computer Theory and Engineering* 2022; 2(2). doi: 10.48550/ARXIV.2201.03391
4. El-Shafai W, Almomani IM, Alkhayer A. Optical bit-plane-based 3D-JST cryptography algorithm with cascaded 2D-FrFT encryption for efficient and secure HEVC communication. *IEEE Access* 2021; 9: 35004–35026. doi: 10.1109/ACCESS.2021.3062403
5. Benrhouma O, Alkhodre AB, AlZahrani A, et al. Using singular value decomposition and chaotic maps for selective encryption of video feeds in smart traffic management. *Applied Sciences* 2022; 12(8): 3917. doi: 10.3390/app12083917
6. Sharma C, Amandeep B, Sobti R, et al. A secured frame selection based video watermarking technique to address quality loss of data: Combining graph based transform, singular valued decomposition, and hyperchaotic encryption. *Security and Communication Networks* 2021; 2021: 1–19. doi: 10.1155/2021/5536170
7. Karmakar J, Pathak A, Nandi D, Mandal MK. Sparse representation based compressive video encryption using hyper-chaos and DNA coding. *Digital Signal Processing* 2021; 117: 103143. doi: 10.1016/j.dsp.2021.103143
8. Helmy M, El-Shafai W, El-Rabaie S, et al. Efficient security framework for reliable wireless 3D video transmission. *Multidimensional Systems and Signal Processing* 2022; 33(1): 181–221. doi: 10.1007/s11045-021-00796-7
9. Yan X, Gilani SZ, Feng M, et al. Self-supervised learning to detect key frames in videos. *Sensors* 2020; 20(23): 6941. doi: 10.3390/s20236941
10. Kelur S, HS RK, K R. Selective area encryption using machine learning technique. In: 2019 Innovations in Power and Advanced Computing Technologies (i-PACT); 22–23 March 2019; Vellore, India. pp. 1–7.
11. Shah RA, Asghar MN, Abdullah S, et al. Effectiveness of crypto-transcoding for H.264/AVC and HEVC video bit-streams. *Multimedia Tools and Applications* 2019; 78(15): 21455–21484. doi: 10.1007/s11042-019-7451-5
12. Peng F, Zhang X, Lin ZX, Long M. A tunable selective encryption scheme for H.265/HEVC based on chroma IPM and coefficient scrambling. *IEEE Transactions on Circuits and Systems for Video Technology* 2020; 30(8): 2765–2780. doi: 10.1109/TCSVT.2019.2924910
13. Xu D. Commutative encryption and data hiding in HEVC video compression. *IEEE Access* 2019; 7: 66028–66041. doi: 10.1109/ACCESS.2019.2916484
14. Lee MK, Jang ES. Start code-based encryption and decryption framework for HEVC. *IEEE Access* 2020; 8: 202910–202918. doi: 10.1109/ACCESS.2020.3036023

15. Long M, Peng F, Li H. Separable reversible data hiding and encryption for HEVC video. *Journal of Real-Time Image Processing* 2018; 14(1): 171–182. doi: 10.1007/s11554-017-0727-y
16. Xu C, Ren W, Yu L, et al. A hierarchical encryption and key management scheme for layered access control on H.264/SVC bitstream in the Internet of Things. *IEEE Internet of Things Journal* 2020; 7(9): 8932–8942. doi: 10.1109/JIOT.2020.2997725
17. Li H, Gu Z, Deng L, et al. A fine-grained video encryption service based on the cloud-fog-local architecture for public and private videos. *Sensors* 2019; 19(24): 5366. doi: 10.3390/s19245366
18. Hassan HE-R, Tahoun M, ElTaweel G. A robust computational DRM framework for protecting multimedia contents using AES and ECC. *Alexandria Engineering Journal* 2020; 59(3): 1275–1286. doi: 10.1016/j.aej.2020.02.020
19. Xu H, Tong X, Wang Z et al. Robust video encryption for H.264 compressed bitstream based on cross-coupled chaotic cipher. *Multimedia System* 2020; 26(4): 363–381. doi: 10.1007/s00530-020-00648-7
20. Huang M, Yang C, Li H, Shen J. Sparse selective encryption for HEVC 4K video using spatial error spread. *Journal of Internet Technology* 2019; 20(5): 1589–1600.
21. Gautier G, FarajAllah M, Hamidouche W, el al. Selective encryption of the versatile video coding standard. *IEEE Access* 2021; 10: 21821–21835. doi: 10.48550/ARXIV.2103.04203
22. Haridas D, Kiran DS, Patel S, et al. Real-Time compressed video encryption: Based on quasigroup on System on Chip (SOC). *SN Computer Science* 2021; 2(5): 408. doi: 10.1007/s42979-021-00793-4
23. Alhassan S, Iddrisu MM, Daabo MI. Perceptual video encryption using orthogonal matrix. *International Journal of Computer Mathematics: Computer Systems Theory* 2019; 4(3–4): 129–139. doi: 10.1080/23799927.2019.1645210.
24. Valli D, Ganesan K. Chaos based video encryption using maps and Ikeda time delay system. *The European Physical Journal Plus* 2017; 132(12): 542. doi: 10.1140/epjp/i2017-11819-7
25. Yasser I, Mohamed MA, Samra AS, Khalifa F. A chaotic-based encryption/decryption framework for secure multimedia communications. *Entropy* 2020; 22(11): 1253. doi: 10.3390/e22111253
26. Huang Q, Zhao X, Li G. Research on the application of video encryption technology based on 7 dimensional CNN hyper chaos. In: 2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA); 10–11 February 2018; Changsha, China. pp. 448–451.
27. Kordov K, Dimitrov G. A new symmetric digital video encryption model. *Cybernetics and Information Technologies* 2021; 21(1): 50–61. doi: 10.2478/cait-2021-0004
28. Li X, Xiao D, Wang QH. Error-free holographic frames encryption with CA pixel-permutation encoding algorithm. *Optics and Lasers in Engineering* 2018; 100: 200–207. doi: 10.1016/j.optlaseng.2017.08.018
29. Waseem HM, Khan M. A new approach to digital content privacy using quantum spin and finite-state machine. *Applied Physics B* 2019; 125(2): 27. doi: 10.1007/s00340-019-7142-y
30. Alattar AM, Al-Regib GI, Al-Semari SA. Improved selective encryption techniques for secure transmission of MPEG video bit-streams. In: Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348); 24–28 October 1999; Kobe, Japan. pp. 256–260.
31. Wen H, Ma L, Liu L, et al. High-quality restoration image encryption using DCT frequency-domain compression coding and chaos. *Scientific Reports* 2022; 12(1): 16523. doi: 10.1038/s41598-022-20145-3
32. Go K, Lee I-G, Kang S, Kim M. Secure video transmission framework for battery-powered video devices. *IEEE Transactions on Dependable and Secure Computing* 2022; 19(1): 275–287. doi: 10.1109/TDSC.2020.2980256
33. Wen H, Liu Z, Lai H, et al. Secure DNA-coding image optical communication using non-degenerate hyperchaos and dynamic secret-key. *Mathematics* 2022; 10(17): 3180. doi: 10.3390/math10173180
34. Farajallah M, Gautier G, Hamidouche W, et al. Selective encryption of the versatile video coding standard. *IEEE Access* 2022; 10: 21821–21835. doi: 10.1109/ACCESS.2022.3149599
35. Faragallah OS, El-Shafai W, Sallam AI, et al. Cybersecurity framework of hybrid watermarking and selective encryption for secure HEVC communication. *Journal of Ambient Intelligence and Humanized Computing* 2022; 13: 1215–1239. doi: 10.1007/s12652-020-02832-z