Journal of Advanced Engineering Trends
ISSN : 2682 - 2091

Vol.43, No.2. July 2024

http://jaet.journals.ekb.eg

AN AUTHENTIC-BASED PRIVACY PRESERVATION SYSTEM FOR SMART E-HEALTHCARE SYSTEMS BASED ON ZERO RUN-LENGTH ENCODING AND DNA CRYPTOGRAPHY

Mona Elamir [(1)], Mai S. Mabrouk[(2)],

Biomedical Engineering Department, Helwan University, Egypt[(1)]

Biomedical Engineering Department, Misr University for Science and Technology, Egypt[(2)]

ARTICLE INFO

Article history:
Received:
Accepted:
Online:

Keywords:
Data Security
DNA cryptography
Lossless data compression
Zero run-length encodings

ABSTRACT

Nowadays, Information security involves protecting such a piece of sensitive information from unauthorized access which includes either inspection, modification, recording, or any disruption or destruction. That's why important strategic resources and large corporations ensure the safety and the privacy of critical data such as customer account details, financial data, or intellectual property. To make sure that the information reaches the intended persons (usually the sender and the receiver), all the weaknesses of security systems must be supported by creating novel algorithms that are based on recent secure technologies like DNA cryptography. This study aimed to propose a crypto-compression system that is based on a hybridization of data compression using zero-Run-Length Encoding (zRLE) and data encryption using DNA cryptography. Such a proposed system reconstructed the secret compressed data with similarity percent 100% (Lossless compression) and zero mean square error (accurate data reconstruction) which resulted in increasing transmission speed for confidential data.

1. Introduction

Data compression is critical in digitally loading information on computer discs and conveying it to communication links. There are two basic compression techniques; lossy (Inexact) or Lossless (Exact) can be used for data compression. To produce the original data, lossless compression may be reversed, whereas lossy compression removes some information or presents minor errors in reversal. For text, lossless compression is appropriate, where each digit is significant, while lossy compression might be reasonable for pictures or speech (The limitation of the frequency spectrum is an example of lossy compression in telephony) [1]. In lossless compression, bits have been reduced by recognizing and removing redundancy in statistics. One of the most lossless compression techniques is "Run Length Encoding (RLE)" which is s considered a simple framework that provides a strong loss-lessness compression for the data values who has the same value [2]. In this algorithm, the repeated run of characters has been coded by using two elements: the first one represents the counter which memorizes how long the data is; the second one recorded the repetitive element that constitutes the data. If a data item (a) occurs at (b) consecutive times in the input data stream, RLE replaced the (b) occurrences with the single pair (ab) as shown in Figure 1.

Revised:1 July , 2022, Accepted:16 September , 2022

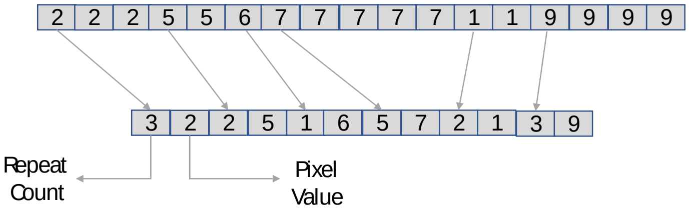

Figure 1: Run Length Encoding algorithm.

Digital revolution communication has some drawbacks like accidentally the secret message could be potentially intercepted and read on its way to the destination. Cryptography here could be a good solution to this problem which involves encrypting such a secret message, hence only the destination person has a key to decrypt it, then privacy will be achieved [3]. Therefore, digital communication should be based on a design that is articulated in two basic operations which are heterogeneous or in some cases conflicting, but it is needed to be applied to the secret file to ensure encrypting efficiency and security. This study handled data encryption using a lossless data compression application.

Different studies handled data security issues as in [4] where hybridization of chaotic maps with DNA cryptography has been used to hide and encrypt both medical images and medical reports. Ponnambalam et al. [5] have proposed a system to convert Electrocardiograph (ECG) signals into QR codes. Mustafa et al. [6] have proposed an encryption algorithm based on the Advanced Encryption Standard algorithm (AES) to improve the security of ECG signals through transmission using three different keys. Triple Data Encryption Standard (Triple-DES) has been implemented in [7] with the aid of the hash function and DNA cryptography base to encrypt different biosignals (ECG, EEG, and EMG) into DNA format to add a security layer for the proposed system.

On the other hand, some studies focused on image compression as in [8] in which a hybrid compression technique is applied on ninety-hand vein images to combine both advantages of lossy and lossless techniques, their goal is to maximize the Compression ratio by preserving the images' details. Their applied technique resulted in a compression ratio reaching 89.56%. in [9] the authors aimed to maximize the compression ratio by preserving images' information, they proved that the maximum accepted ratio for each image is chosen by three experts. Their last enhancement stage is done by isolating the region of interest (ROI) in the image then applying the compression procedure by the experts gave promising results.

II. Methodology

In this study, a crypto compression scheme has been proposed using the hybridization of a novel updated version of RLE called zero-RLE is applied for performing lossless data compression on the medical data (biosignals, images, and videos) with DNA encoding rules. the compressed data seemed like an encrypted form which is required in data security then finally the compressed data is encoded into DNA format as shown in Figure 2.

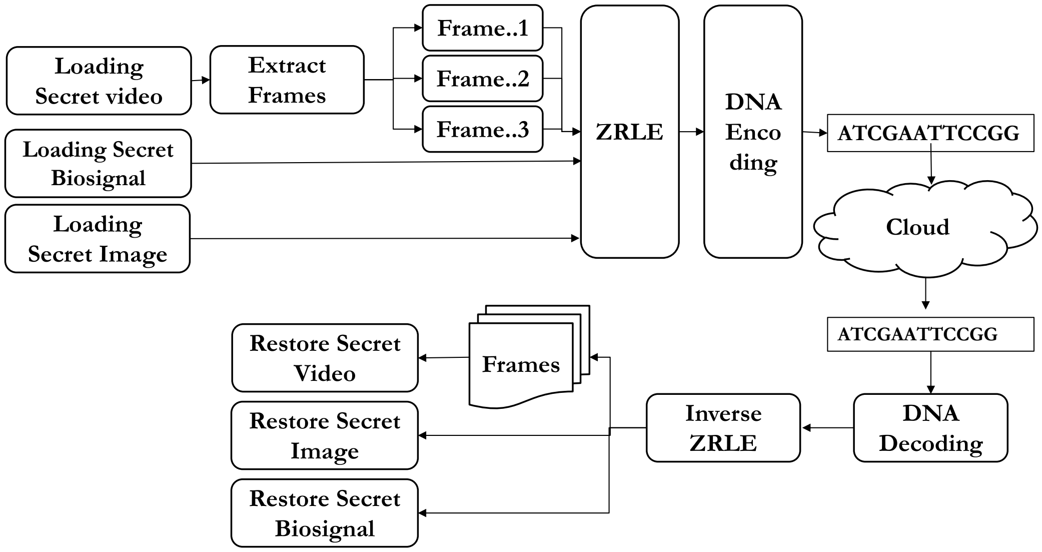

Figure 2: Block diagram of the proposed compression-encryption scheme.

A. Dataset

In this study, three types of data are utilized in Appling the prop any types of c

Table 1: The dataset used in this study

Data Type	Data description
Medical Images	Chest X-ray images have been tested for covid-19 [10]
Biosignals	Electrocardiograph (ECG) signals downloaded from the MIT-BIH Arrhythmia database [11]. Electroencephalograph (EEG) signals down were loaded from the DEAP dataset for EEG signals recorded at different emotions [12]. Electromyograph (EMG) downloaded from EMG data for gestures Data Set These are files of raw EMG data recorded by MYO Thalamic bracelet In case hand clenched in a fist [13].
Medical Videos	Twenty eyes videos of 24 subjects were included in the study about the Field of View (FOV) of eyes. Each grayscale image in the sequence was composed of 768- × 768-pixel images [14].

B. Zero Run-length Encoding (zRLE)

Zero-Run Length Encoding is an updated version of RLE which originated in [15] for lossless compressing for implantable optogenetic visual prostheses. zRLE involves checking the pixel value and comparing it with specific ecif threshold; if the pixel value is higher than the threshold value hence keeping the pixel value as it is, if the pixel value is less than the threshold value hence rep replace with zero, the threshold value for each type of data is calculated from minimum peaks in the histogram. Finally, check all zeros: (For non-repeated zero replaced with one- For repeated zero, replaced with (count, zero)). The flowchart of zRLE is shown in Figure 3.

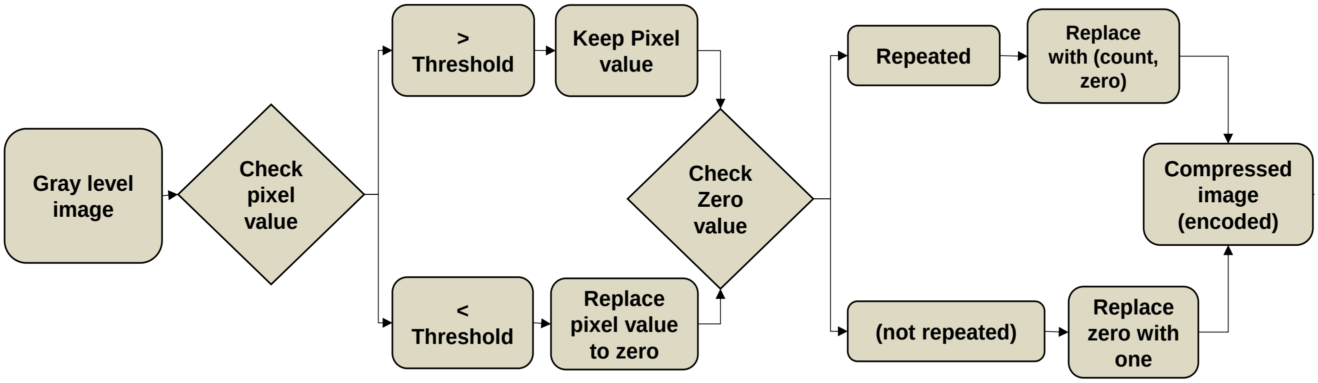

Figure 3: Flowchart for zRLE.

C. DNA Encoding

It is a novel technology that involves utilizing DNA molecules in saving data it. In digital applications converting binary data into a DNA sequence that's adding a new security level for data encryption [16]. Eight probability encoding rules may be used for converting binary data into DNA sequence as shown in Table 2. In this study, Rule-3 is used.

Table 2: DNA Encoding rules

	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rule 8
A	**00**	**00**	**01**	**01**	**10**	**10**	**11**	**11**
C	**10**	**01**	**11**	**00**	**11**	**00**	**10**	**01**
G	**01**	**10**	**00**	**11**	**00**	**11**	**01**	**10**
T	**11**	**11**	**10**	**10**	**01**	**01**	**00**	**00**

D. System Evaluation Metrics

Several evaluation metrics can be used to evaluate the privacy, robustness, and efficiency of the proposed algorithm. In this study, seven evaluation metrics [17] are used:

a. Encryption/ Decryption time: represents the time of encryption or decryption.

b. Histogram: represents the frequency distribution of the pixel intensity values. The histogram of the encrypted image should be completely differed from the histogram of the original image.

$$Compression\ Ratio = \frac{un\ compressed\ Size}{Compressed\ Size}$$

c. Correlation coefficient: measures the strength of the linear relationship between variables, it can be calculated from:

$$\boldsymbol{r_{\alpha\beta}} = (\frac{\boldsymbol{cov(\alpha,\beta)}}{(\sqrt{\boldsymbol{VAR(\alpha)}})(\sqrt{\boldsymbol{VAR(\beta)}})}) \qquad (1)$$

$$\boldsymbol{VAR(\alpha) = \frac{1}{N}\sum_{i}^{N}\left(\alpha_i - E(x)\right)^2} \qquad (2)$$

$$\boldsymbol{cov(\alpha,\beta) = \frac{1}{N}\sum_{i=1}^{N}(\alpha_i - E(\alpha)) * (\beta_i - E(\beta))} \qquad (3)$$

Where: (α, β) are the original and reconstructed data. (i): pixel which calculation is done. E(x): Data Expectation. $cov(\alpha,\beta)$: is the covariance between α and β. $VAR(\alpha)$ gives the value of variance at pixel value α, N: is the total number of pixels.

d. Mean Squared Error (MSE) analysis: Mean Squared Error (MSE) has been used to check the diffusion characteristic of an image of the cryptosystem. It indicates that the relationship is too complex to be predicted easily.

$$\boldsymbol{MSE = \left(\frac{1}{WH}\right) * \sum_{i=1}^{W}\sum_{j=1}^{H}[I(i,j) - K(i,j)]{}^\wedge 2} \qquad (4)$$

Where I and K refer to the pixel values of both original and encrypted images. (i,j) refer to the pixel location. (W and H) refer to the image dimensions.

e. Unified Average Changing Intensity (UACI):
UACI is the difference between the average intensity between the plain and encrypted images. The theoretical value for UACI is about 33%. UACI can be calculated from:

$$\boldsymbol{UACI} = \sum_{i,j} \frac{|\boldsymbol{C1}(\boldsymbol{i},\boldsymbol{j}) - \boldsymbol{C2}(\boldsymbol{i},\boldsymbol{j})|}{\boldsymbol{F} * \boldsymbol{T}} * \boldsymbol{100}\% \qquad (5)$$

(C1, C2): are the original and reconstructed data. (T): represents the whole number of the pixels in the encrypted image, (F): represents the largest supported pixel value (it will be 255).

f. Structure Similarity Index Measurement (SSIM):
SSIM measures the similarity between two images. SSIM can be calculated from:

$$\boldsymbol{SSIM}(\boldsymbol{x},\boldsymbol{y}) = \frac{(2\boldsymbol{\mu_x}\boldsymbol{\mu_y} + \boldsymbol{C1})(2\boldsymbol{\sigma_{xy}} + \boldsymbol{C2})}{(\boldsymbol{\mu_x^2} + \boldsymbol{\mu_y^2} + \boldsymbol{C1})(\boldsymbol{\sigma_x^2} + \boldsymbol{\sigma_y^2} + \boldsymbol{C2})} \qquad (6)$$

Where: (x, y): is the original and reconstructed data. (μ_x , μ_y) is the average of x, y. (σ_x, σ_y) is the variance of x, y. (σ_{xy}) is the covariance of x, y. (C1, C2) are two constants to stabilize the division with a weak denominator.

i. Compression Ratio: it represents the ratio between compressed data size to original data size.

2. Results

The goal of this study is to propose a crypto-compression system to perform both encryptions for data for security and compression for more transmission speed in a small time to enable medical teamwork in handling medical data safely when transmitting through public channels. Zero Run-length encodings, a lossless compression algorithm, are used in this study to reduce data size with the aid of DNA encoding rules for data encryption. This experiment is implemented on Windows 10 using MATLAB (R2018a) on a personal laptop with the following specifications: CPU 2.7 GHz Core i7, and 8 GB RAM.

Different evaluation metrics have been calculated starting with the histogram of original data and decompressed one which is shown in Fig. 4 where the histogram of the original data was like the histogram of the decompressed one. This proposed security scheme is achieved a satisfactory result and a good performance with values reached to the ideal values (100% similarity and zero MSE) which have appeared after applying these evaluated metrics. The correlation between the original data and the restored one indicated its correlated relation as they were identical.

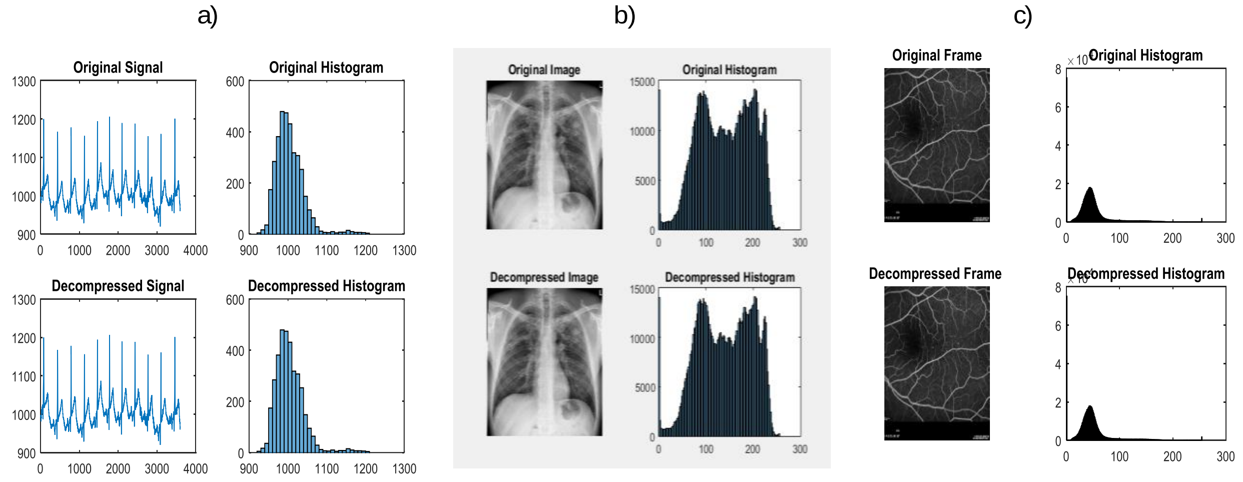

Figure 4: The histogram of the original data (a) for biosignal (1D data), b) for image (2D data), c) for Videoframe data) and the decompressed one using zRLE

Table 3 showed the statistical evaluation metrics achieved by the proposed cryptosystem with each type of data. In Table 3, Encryption/ Decryption time for biosignal and images were very small which indicated its ability to real-time applicability. SSIM was 100% which means that the input historical with the output that proves lossless compression, MSE is also equal to zero due to lossless compression with zero error. The compression ratio was differentiated between (10% to 25%).

Table 3: Experimental result for the proposed crypto-compression system

	Medical Images	Biosignals			Medical Videos
		ECG	EEG	EMG	
Encryption time (s)	11.6 s	10 s	12.123 s	15.2 s	402 s
Decryption time (s)	7.5 s	8.78 s	11.04 s	12.8 s	250 s
SSIM	100 %	100 %	100 %	100 %	100 %
MSE	0	0	0	0	0
UACI	33.3	33.34	33.5	33.38	34
Correlation Coefficient between original and reconstructed data	1	1	1	1	1 (Completely correlated)
Compression ratio	16 %	25 %	13 %	10 %	12 %

3. Discussion

Proposing a crypto-compression system requires the essential specification to perform its goal. In this study, zRLE proved its efficiency in performing lossless compression for data without a loss (zero MSE), so we must compare our results with other studies which focused on using RLE in the same application or in using it in another application. Firstly, we have compared our result in medical image compression and encryption with other studies that have the same goal on images as in Table 4:

Table 4: The Comparing of the proposed crypto-compression system of image encryption with other studies

Author	Methodology	Data specs	Compression ratio
Sh. Khaleel [18]	Fuzzy Neural Network Clustering with RLE	Tree gray img. 128*128	Tree time.: 13.86 %
T. Agung [19]	Watermarking scheme with RLE	fetal-lens-1b.bmp 640x480	1.32 %
Our study of image	zRLE with DNA encoding	Medical in. 784*1024	16 %

Other studies have focused on 1D data encryption like audio signal, and biosignals, here we have compared these studies with other system applications on biosignals. The result of the proposed system proved that our proposed methodology could perform both encryption and compression in a small time with zero MSE which indicated the data has been reconstructed exactly through the process. Also, the compression ratio concerning others is accepted and achieved in a good time as shown in Table 5

Table 5: Comparing the proposed crypto-compression system of biosignal encryption with other studies

	Methodology	Data specs	Compression ration
S. Akhter,2010 [20]	(RLE) to compress (DCT) coefficients of ECG.	MIT-BIH Arrhythmia	14.87 %
B. Carpentieri, 2018 [21]	RLE +3DES	1D data: 3229KB	21 %
Our study in ECG biosignals	RLE+DNA Encoding	ECG biosignals from MIT-BIH Arrhythmia	25 %

4. Conclusions

Medical data is very sensitive data as it contains patients' data besides medical reports, biosignals, medical images, and medical videos. These data can't be sent online without encryption as it can be hacked, alternated, modified, or deleted. Although there are different methodologies for data compression and encryption, Zero-Run length encoding performs both fast and good compression ratios for data then it is supported with DNA encoding rules to enhance security levels through data transmission. The proposed scheme is very fast and accurate with zero error and 100% similarity percent that can be applied to real-time data compression and encryption. In the future, we will support such a proposed system with one of the cryptographic algorithms like AES to enlarge keyspace and hence increase the security level for the proposed system.

Conflict of Interest

The authors declare no conflict of interest.

Abbreviation and symbols

Symbol	**Abbreviation**
AES	Advanced Encryption Standard
DCT	Discrete Cosine Transform
DES	Digital Encryption Standard
DNA	Deoxyribonucleic acid
ECG	electrocardiograph
EEG	Electroencephalograph
EMG	Electromyograph
FOV	Field of View
MSE	Mean Squared Error
RLE	Run Length Encoding
SSIM	Structure Similarity Index Measurement
UACI	Unified Average Changing Intensity
zRLE	Zero Run-Length Encoding

References

[1] K. Sayood, *Introduction to data compression*: Morgan Kaufmann, 2017.

[2] H. Ren, et al., "On state-space compression and state reachability retrieval of Petri nets," *Advances in Mechanical Engineering,* Vol. 11(2) 1–15, 10.1177/1687814019825962, 2019.

[3] P. Pandhare, et al. , "Cryptography Technique for Information Security " *International Journal of Advanced Research in Computer and Communication Engineering* vol. Vol. 8, p. 321-325, April 2019.

[4] M. Elamir, et al., "Hybrid image encryption scheme for secure E-health systems," *Network Modeling Analysis in Health Informatics and Bioinformatics,* vol. 10, pp. 1-8, 2021.

[5] P. Mathivanan, et al., "QR code–based ECG signal encryption/decryption algorithm," *Cryptologia,* vol. 43, pp. 233-253, 2019.

[6] M. Hameed, et al., "Comparative study of several operation modes of AES algorithm for encryption ECG biomedical signal," *International Journal of Electrical & Computer Engineering (2088-8708),* Vol. 9, No. 6, December 2019.

[7] M. Elamir, et al., "An E-health System for Encrypting Biosignals Using Triple-DES and Hash Function," in *2021 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, 2021, pp. 11-14.

[8] M. Saad and A. Kandil, " A hybrid compression technique for segmented hand veins using quad tree decomposition," in *2012 Cairo International Biomedical Engineering Conference (CIBEC)*, 2012, pp. 162-165.

[9] M. Saad and A. Kandil, "Improved ROI Algorithm for Compressing Medical Images," in *BIODEVICES*, 2013, pp. 184-189.

[10] J. Cohen, et al., "Covid-19 image data collection: Prospective predictions are the future," *arXiv preprint arXiv:2006.11988,* 2020.

[11] A. Goldberger*, et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *circulation,* vol. 101, pp. e215-e220, 2000.

[12] S. Koelstra, *et al.*, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing,* vol. 3, pp. 18-31, 2011.

[13] S. Lobov, et al., "Latent factors limiting the performance of sEMG-interfaces," *Sensors,* vol. 18, p. 1122, 2018.

[14] H. Rabbani, et al., "Fully automatic segmentation of fluorescein leakage in subjects with diabetic macular edema," *Investigative ophthalmology & visual science,* vol. 56, pp. 1482-1492, 2015.

[15] Z. Hou, et al., "A scalable data transmission scheme for implantable optogenetic visual prostheses," *Journal of Neural Engineering,* vol. 17, p. 055001, 2020.

[16] A. Das, et al., "Data security with DNA cryptography," in *Transactions on Engineering Technologies*, ed: Springer, 2021, pp. 159-173.

[17] M. Pedersen and J. Hardeberg, "Full-reference image quality metrics: Classification and evaluation," *Foundations and Trends® in Computer Graphics and Vision,* vol. 7, pp. 1-80, 2012.

[18] S. Khaleel and B. Khaleel, "Image Compression Based on Artificial Intelligent Techniques," *AL-Rafidain Journal of Computer Sciences and Mathematics,* vol. 6, pp. 75-109, 2009.

[19] T. BW and F. Permana, "Medical image watermarking with tamper detection and recovery using reversible watermarking with LSB modification and run length encoding (RLE) compression," in *2012 IEEE International Conference on Communication, Networks and Satellite (ComNetSat)*, 2012, pp. 167-171.

[20] S. Akhter and M. Haque, "ECG comptression using run length encoding," in *2010 18th European Signal Processing Conference*, 2010, pp. 1645-1649.

[21] B. Carpentieri, "Efficient compression and encryption for digital data transmission," *Security and Communication Networks*, 9591768, 10.1155/2018/9591768, Volume 2018.