






## Strengthening Cryptographic Keys Using User's Voice Biometric

Ahmed A. Abbass<sup>1</sup>, Hussein L. Hussein<sup>2\*</sup>, Sinan A. Naji<sup>1</sup>, Jasim H. Lafta<sup>3</sup>, Robert Tornai<sup>4</sup>

<sup>1</sup> College of Business Informatics, University of Information Technology and Communications, Baghdad 10001, Iraq

<sup>2</sup> Department of Computer Science, College of Education for Pure Sciences, University of Baghdad, Baghdad 10001, Iraq

<sup>3</sup> Department of Computer Engineering Techniques, Al-Bayan University, Baghdad 10001, Iraq

<sup>4</sup> Department of Data Science and Visualization, Faculty of Informatics, University of Debrecen, Debrecen 4000, Hungary

Corresponding Author Email: [hussein.l.h@ihcoedu.uobaghdad.edu.iq](mailto:hussein.l.h@ihcoedu.uobaghdad.edu.iq)

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijssse.150516>

### ABSTRACT

**Received:** 6 February 2025

**Revised:** 15 March 2025

**Accepted:** 24 April 2025

**Available online:** 31 May 2025

#### **Keywords:**

*cryptography, encryption, AES, histograms, thresholding, voice biometric*

Nowadays, the most trustworthy tool to protect our information is based on cryptography. To enhance cryptography, it is highly important to select a robust cryptographic key that implies a high degree of uniqueness and randomness. In general, users lack proficiency in generating strong cryptographic keys. Furthermore, they often struggle with remembering how to generate strong cryptographic keys. This paper introduces an interesting method to abridge and boost the key generation step. The proposed method makes use of audio files that contain the user's voice biometric. A variable histogram-thresholding, which is calculated for each audio file distinctly, was used to get approximation peaks in the audio file and then to extract the discriminating features that ensure the uniqueness. The features will then be normalized and converted from a decimal to binary system to generate two 128-bit subkeys. Finally, to generate a robust AES-128-bit encryption key, these two subkeys are shifted, mixed, and XORed to enhance randomness. A dataset of more than 900 audio files is used. Testing and evaluation demonstrate how robust the suggested system is for helping users generate robust cryptographic keys.

## 1. INTRODUCTION

Internet cyberspace, as an openly available resource, allows users to acquire and exchange different types of information. This situation, plus the recent advancements in the Internet of Things (IoT), raises many privacy and security issues for individuals and organizations that have to deal with cyberattack challenges [1, 2]. Fortunately, cryptography and steganography are perfect techniques to protect our data [3]. Most cryptography techniques use encryption keys and special algorithms to encrypt data [4-6]. This makes sure that without the necessary key, attackers cannot decrypt our ciphered data, even if they manage to get their hands on it. Generally, the strength of the cryptography process requires a strong encryption key [7-9]. To strengthen encryption keys, many recent systems take advantage of biometric mechanisms [10, 11]. User biometrics offer reliable methods for user authentication. Numerous feature extraction techniques have been proposed to extract some kind of discriminating characteristics from voice, iris, fingerprint, face, etc. [12, 13]. These characteristics can be used for generating a secure encryption key [14-16]. In practice, user voice patterns are efficient biometric and preferred by the general population for two reasons [17-19]: first, people do not like to use things that are used by others. For instance, the same biometric reader or scanner is used to take samples from a group of individuals. Generally, capturing a user's voice sample typically doesn't require any physical interaction with other devices. Second,

voice samples offer excellent randomness, which is important when creating encryption keys. This ensures that the generated key is robust and resistant to brute-force attacks.

In this study [20], a new technique for strengthening key generation is proposed that combines the user's voice, histogram generation, peak thresholding, and finally mixing and permutation. The main idea is to extract control features from the human voice. Then, these control features will be passed as control points for creating two subkeys. To enhance the randomization and confusion that generates a strong key, the first subkey would be shifted and then mixed by XOR operation with the second key. A dataset of more than 900 audio files had been collected from the study [20].

The rest of this study is divided as follows: Section 2 presents a brief introduction to the AES-128 encryption algorithm. The proposed system is described in Section 3. Testing and evolution are presented in Section 4. Finally, Section 5 presents the conclusions and future work directions.

## 2. THE AES ENCRYPTION ALGORITHM

The Advanced Encryption Standard (AES) is an interesting encryption method that has been practically applied to protect different types of data in many applications [5, 21]. It is classified as a modern symmetric-key block cipher that eclipses the previous techniques. The AES performs perfectly in three criteria: security, cost, and implementation [5, 21].

The strength of AES comes from operating at various levels of data units: bit, byte, word, state, and block along with four types of transformations: substitution, permutation, mixing, and key-adding. Many different AES versions are proposed such as AES-128, 192, and 256-bit encryption algorithms. The key size, which can be 128, 192, or 256 bits, depends on the number of rounds as shown in Table 1 [5, 21]. As the structure of AES-128 is publicly available, it is highly important to generate a strong encryption key to encrypt our data. The next section describes the proposed system to generate a strong AES-128 bit encryption key.

**Table 1.** Advanced Encryption Standard (AES) possible combinations

Key Length	No. of Rounds	Possible Combinations
128 Bits	10	$3.4 \times 10^{38}$
192 Bits	12	$6.2 \times 10^{57}$
256 Bits	14	$1.1 \times 10^{77}$

### 3. THE PROPOSED SYSTEM

This section describes the main steps to generate a strong AES-128 encryption key with high randomness and uniqueness. The main idea of the proposed method is to generate the encryption key from voice biometrics. The proposed system implies four main steps: histogram generation for audio file, thresholding peaks in voice, extracting features, and encryption key generation.

#### 3.1 Histogram generating for audio file

The proposed method begins with a pre-processing step that reads the audio file contents. This stage will first read the chunk of the header bytes in the audio file until voice data is reached. The audio data will be used as input to the histogram generation step. A histogram is some kind of plot that shows the number of data points with a specific range of values in the dataset (i.e., count of occurrences). Initially, the minimum and maximum data values are determined to compute the range.

The range was divided into consecutive, non-overlapping intervals (i.e., bins) that represent the entire range. Then, count how many data values belong to each interval. Generally, the dataset and the range of a histogram are normalized to 1, and then passed to peaks-thresholding step. Figure 1 shows the histogram of a sample audio file.

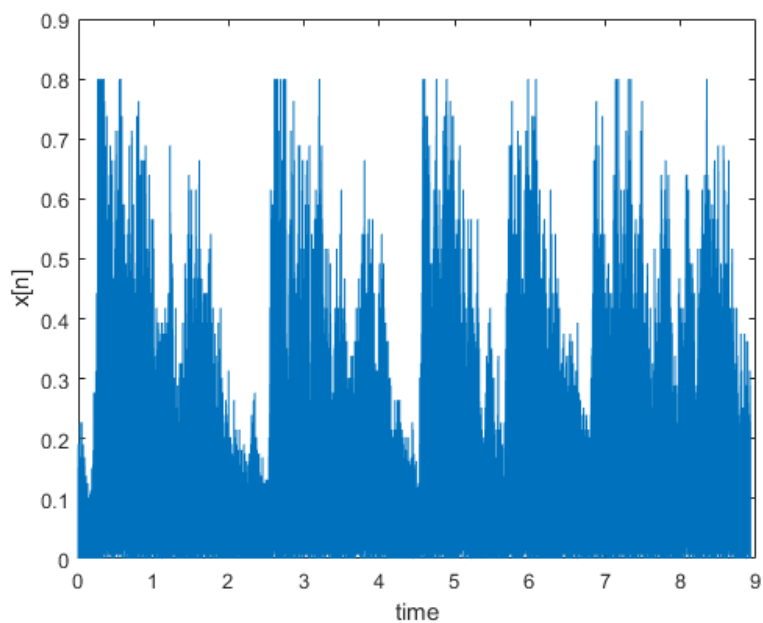
#### 3.2 Thresholding peaks in voice

Thresholding in computer science is a fast and interesting method that is highly used to deal with and analyze different data types. Thresholding can be applied to perform many tasks, such as segmentation, color adjustment, audio data processing, etc. Generally, the threshold  $T$  is a value calculated according to the task to be done. But, in some cases, the task at hand may require multiple thresholds. Furthermore, the threshold may be a constant value referred to as global-thresholding or it may vary over the input data and be referred to as variable-thresholding. In this study, a peak-thresholding step would be applied on the histogram rather than the actual data, which shows the peaks and valleys between the voice data. We propose to use variable-thresholding, which is calculated for each audio file distinctly (i.e., each individual’s voice) as follows:

1. Choose an initial threshold value  $T$ , where the mean of peaks  $\mu_0$  is used as the initial value  $T$ .
2. Split the audio file data using  $T$  to generate two sets of audio data:  $G_1$  involving audio data with levels  $> T$  and  $G_2$  involving audio data with levels  $\leq T$ .
3. Compute the mean levels of values in  $G_1$  as  $\mu_1$  and  $G_2$  as  $\mu_2$ .
4. Compute a new threshold value  $T$ :

$$T = \frac{\mu_1 + \mu_2}{2} \tag{1}$$

5. Repeat steps (2-4) until the difference in  $T$  in successive iterations is less than a predefined value  $\epsilon$ .  
Practically, this algorithm shows excellent results for finding thresholds  $T$  for each voice biometric data distinctly.



**Figure 1.** Histogram of a sample audio file

### 3.3 Extracting features

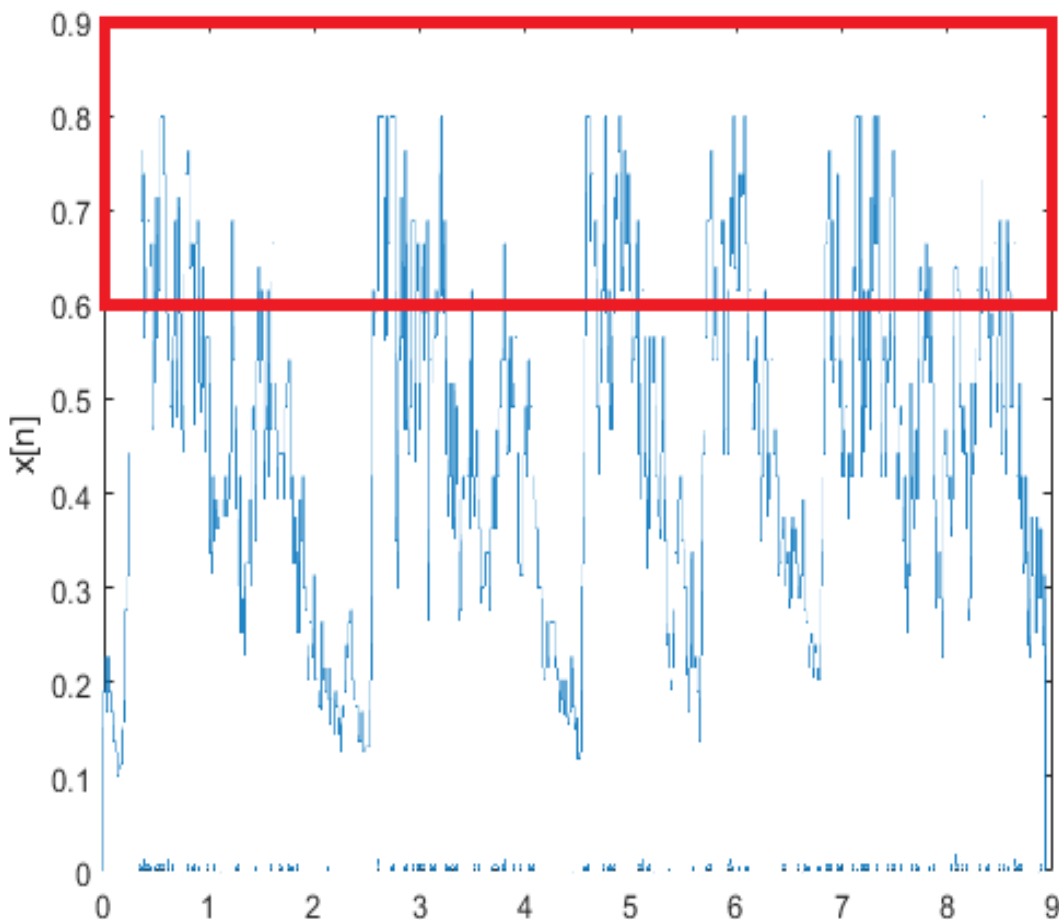
The peaks-thresholding aims at determining the discriminating features of the peaks in an audio file. This is accomplished by extracting the highest peaks, as shown in Figure 2, and dividing the thresholded peaks by  $N$ , where  $N$  is the sample's size. Then, multiply the peaks output with  $1E+8$  to get a sequence of 8-digit integers (i.e.,  $w_1, w_2, w_3, w_4$ ) as shown in Figure 3(a) and (b).

### 3.4 Encryption key generation

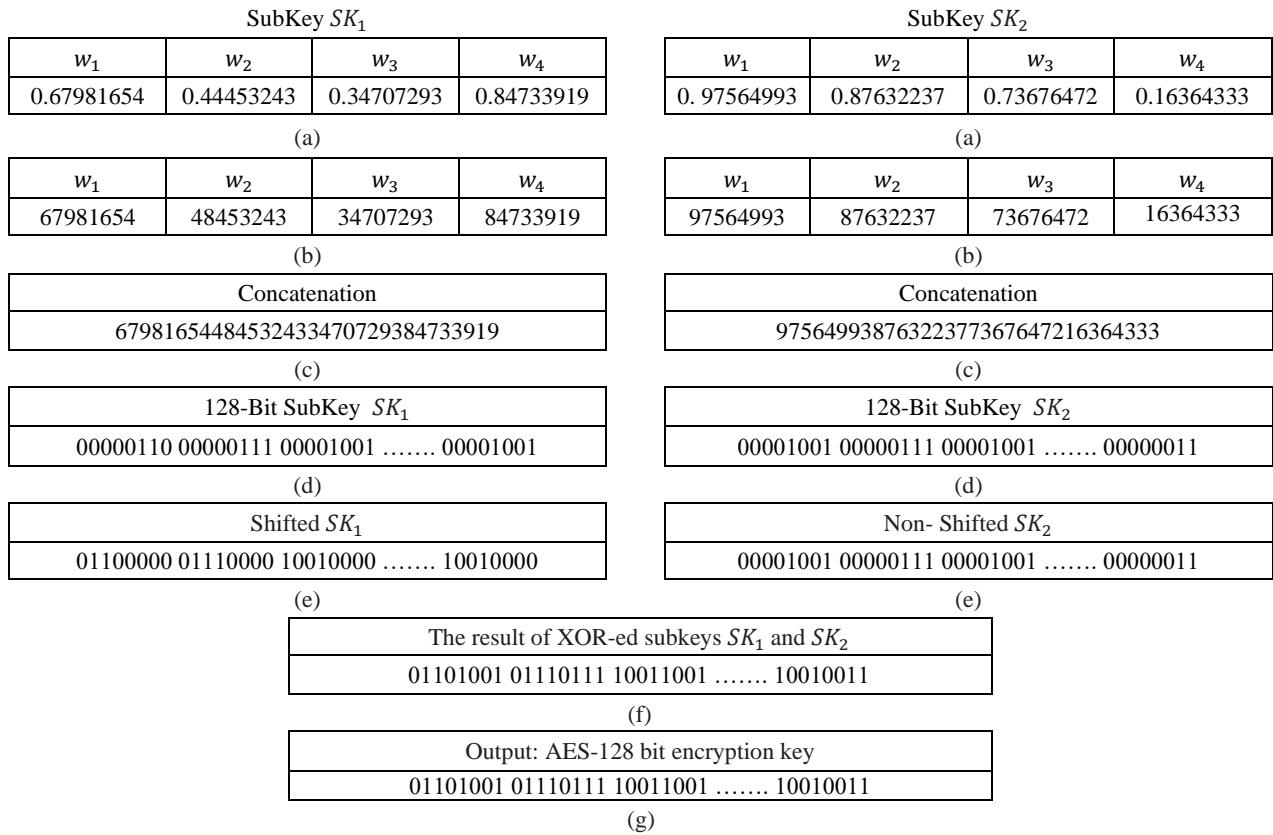
In general, users lack proficiency for generating strong cryptographic keys. Furthermore, they are usually tackling how to remember how to create strong cryptographic keys. This section illustrates how the encryption key is generated from the user's voice biometric with high randomness and uniqueness. The main idea of the proposed method is to generate two subkeys based on the audio voice file in order to ensure the uniqueness. Practically, an important component in most block ciphers is the exclusive-or operation (i.e. XOR). So, these two subkeys are shifted, mixed, and XORed to enhance the randomness. As mentioned in the previous step (see Section 3.3), four 8-digit integers (i.e.,  $w_1, w_2, w_3, w_4$ ) would be concatenated to produce a 32-digit integer. Then, convert each two decimal digits into its binary representation

to produce a 128-bit subkey. To overcome the weakness of the 8-bit conversion step that usually produces 0's more than 1's in 8-bit representation and to get a high randomness (e.g., the binary representation of number 21 is 00010101 that involves many 0's), we propose that each two subkeys  $SK_1$  and  $SK_2$  are used to produce a robust AES-128 encryption key as follows: the subkey  $SK_1$  would be shifted 4 digits to the left and then XOR-ed with the subkey  $SK_2$  as shown in Figure 3(f). The proposed permutation by the shift transformation exchanges bytes without permuting the bits inside the bytes, while the XOR carries out an interbyte transformation that changes the bits inside a byte based on the bits inside the other byte.

As shown in Figure 3, a 128-bit encryption key is generated based on the individual's voice biometric (i.e., audio file). Figure 3(a) shows a sample of voice peaks-thresholding results. Figure 3(b) shows the peaks multiplied by  $1E+8$  to produce 8-digit integers. Figure 3(c) shows the concatenation step to generate a 32-digit integer. Figure 3(d) shows the results of conversion into binary representation Figure 3(e) shows the 4 bits shifted of subkey  $SK_1$ . The result of XOR operation of the two subkeys is shown in Figure 3(f). Finally, Figure 3(g) shows the generation of a robust AES-128 encryption key. Testing and evaluation of sample keys that are generated by the proposed system are presented in the next section.



**Figure 2.** Extracting the highest peaks using thresholding



**Figure 3.** The key steps of the proposed method; (a) Peaks-thresholding; (b) Converted into 8-digit integers; (c) Concatenation step; (d) The 128-bit binary representation; (e) The 4 bits rotation results; (f) The result of XORing the two subkeys  $SK_1$  and  $SK_2$ ; (g) The generation of a robust AES-128 encryption key

#### 4. TESTING AND EVALUATION

Testing and evaluation are crucial steps for security evaluation of a cryptographic system. Generally, randomness constitutes a necessary part of many tests. This section is devoted to verifying the strength of the keys generated by the proposed system using different testing and evaluation metrics. According to the NIST suite, various statistical tests can be applied. In this study, four empirical statistical tests were performed on the 128-bit sample keys, these are: the Entropy Test, Frequency Test, Frequency Mono Test, and Runs Test. Table 2 shows the detailed explanation of these tests along with success/failure threshold values. In these tests, the selected statistics were calculated for a dataset composed of more than 900 audio files that had been collected by Ref. [20]. Table 3 shows the testing and evaluation results of five sample AES-128 keys generated by the proposed system. The experimental results reveal that the generated AES-128 keys are robust and passed the success/failure measures, as shown in Table 3.

This innovative approach differs significantly from other works. Monroe et al. [14] proposed a key generation system

based on voice biometrics utilizing the user's spoken password. The main drawback of the spoken password method is that an attacker would have the opportunity to hear/record the user's voice while speaking their password. This makes the system vulnerable to a risk. Conversely, the proposed method overcomes this drawback, making it more reliable for key generation. The proposed system does not directly use the user's voice/password. Instead, the voice-biometric file is processed through highly secure steps to generate the key. Other works, such as in Refs. [8, 9, 11, 15], have explored the feasibility of biometric-based key generation. The main drawback of such systems is the use of static biometrics (such as fingerprints, irises), which are highly suitable for controlling access gates but are not preferred for key generation, as using static samples with systematic algorithms would result in static keys. Furthermore, such biometrics are exposed to a similar risk, as the iris can be captured by a hidden camera, and fingerprints left on surfaces can be lifted hours later. Conversely, the proposed method overcomes this drawback by using its applicability to use different voice files for the same user, which would generate multiple derived keys.

**Table 2.** The empirical statistical tests

Test	Meaning	Success Measure	Explanation
Entropy Test	Randomness of data	close to 1	More robust
Frequency Test	The occurrences of a specified symbol	$0 < X^2 < 3.84$	Success
Frequency MONO Test	The number of symbols in a key are nearly equal.	$P\_value > 0.01$	Success
Runs Test	Investigate the number of runs in a sequence	$P\_value > 0.05$	Success

**Table 3.** Comparing the randomness of keys generated from different audio files

Keys Generated by the Proposed System	Freq. Test	Freq. Mono Test	Runs Test	Entropy Test
0001000010111001001001010100010100010000011100001 ..... 0111011101111	3.781250e+00	5.576726e-01	4.211455e-01	9.785845e-01
10001101101011101100111100010110000 10000011100001 ..... 0100010001100	1.531250e+00	6.399940e-01	1.0	9.913533e-01
11001010011101010110010110111001011100001100001 ..... 111101001111100	3.125000e-02	8.596838e-01	6.239325e-02	9.998239e-01
100110001001101001100101110110011000110110101110 ..... 00110011001100	3.125000e-02	8.596838e-01	4.232813e-01	9.998239e-01
0011000001010101110111010001101101110000100000 ..... 100100010010011	2.812500e-01	7.594629e-01	6.752074e-01	9.984144e-01

## 5. CONCLUSION

The structure of many cryptography algorithms, including AES-128, is publicly available and open for users. So, the strength of the cryptography must be guaranteed by the cryptography key. Therefore, it is crucial to use a key in such a way that will be difficult for the adversary to break. On the other hand, using weak random values in key generations can cause a leakage in the system, and hence an adversary can gain the ability to break the whole cryptosystem. In general, users are not skilled at creating secure cryptographic keys. In this study, a new method is proposed to generate encryption keys for the AES-128 encryption algorithms. The proposed system is based on the adoption of several techniques in one integrated system to get a combination of randomization and uniqueness to enhance the key generation step. The suggested system extracts discriminatory features from the audio files that contain the user's voice. The features will then be normalized to generate two 128-bit subkeys. Finally, to generate a robust AES-128 encryption key, these two subkeys are rotated, mixed, and XORed to enhance the randomness.

Testing and evaluation showed that the threshold values are not exceeded by any of the randomization test findings. Thus, the generated keys have passed the success/failure measure and satisfy the necessary requirements for unpredictability and randomness so they can be used successfully for AES-128 cipher systems.

Future research directions will concentrate on extracting the best peaks from audio recordings using machine learning and artificial intelligence techniques. The key generation task will be greatly accelerated by the use of parallel processors, which will increase the applicability of the proposed system.

## REFERENCES

- [1] Tang, Q., Du, F. (2021). Internet of Things Security: Principles and Practice. Springer. <https://doi.org/10.1007/978-981-15-9942-2>
- [2] Abbass, A.A., Hussein, H.L., Kaabi, J.H.L., Tornai, R. (2022). American standard code for information interchange mapping technique for text hiding in the RGB and gray images. International Journal of Electrical and Computer Engineering (IJECE), 12(3): 2812-2817. <https://doi.org/10.11591/ijece.v12i3.pp2812-2817>
- [3] Sulavko, A., Panfilova, I., Inivatov, D., Lozhnikov, P., Vulfin, A., Samotuga, A. (2025). Biometric-based key generation and user authentication using voice password images and neural fuzzy extractor. Applied System

- Innovation, 8(1): 13. <https://doi.org/10.3390/asi8010013>
- [4] Odhaib, A.R. (2021). Using audio noise for generating random key stream. IOP Conference Series: Materials Science and Engineering, 1090(1): 012136. <https://doi.org/10.1088/1757-899X/1090/1/012136>
- [5] Forouzan, B.A. (2011). Cryptography & Network Security. McGraw-Hill, Inc.
- [6] Faisal, M., Ali, I., Khan, M.S., Kim, J., Kim, S.M. (2020). Cyber security and key management issues for Internet of Things: Techniques, requirements, and challenges. Complexity, 2020(1): 6619498. <https://doi.org/10.1155/2020/6619498>
- [7] Bhaya, W., Abbass, A.A. (2015). Mersenne prime number generating using cubic spline to be used RSA algorithm. Journal of Next Generation Information Technology, 6(1): 1.
- [8] Panchal, G., Samanta, D. (2018). A novel approach to fingerprint biometric-based cryptographic key generation and its applications to storage security. Computers & Electrical Engineering, 69: 461-478. <https://doi.org/10.1016/j.compeleceng.2018.01.028>
- [9] Al-Rifae, Z.I., Ismaeel, T.Z., Abood, S.I. (2024). Cryptography based on fingerprint bio metrics. Journal of Internet Services and Information Security (JISIS), 14(4): 401-417. <https://doi.org/10.58346/JISIS.2024.I4.025>
- [10] El-Shafai, W., El-Mesady, A., Kamal, F.M. (2024). Enhancing biometric authentication security through the novel integration of graph theory encryption and chaotic logistic mapping. Multimedia Tools and Applications, 84: 16909-16943. <https://doi.org/10.1007/s11042-024-19693-9>
- [11] Patel, A.K., Paul, D., Giri, S., Chaudhary, S., Gautam, B. (2024). Gradient-based facial encoding for key generation to encrypt and decrypt multimedia data. arXiv preprint arXiv:2412.06927. <https://doi.org/10.48550/arXiv.2412.06927>
- [12] Naji, S.A., Tornai, R., Lafta, J.H., Hussein, H.L. (2020). Iris recognition using localized Zernike features with partial iris pattern. New Trends in Information and Communications Technology Applications. NTICT 2020. Communications in Computer and Information Science, vol 1183. Springer, Cham. [https://doi.org/10.1007/978-3-030-55340-1\\_16](https://doi.org/10.1007/978-3-030-55340-1_16)
- [13] Jaleel, H.Q., Stephan, J.J., Naji, S.A. (2022). Gender identification from speech recognition using machine learning techniques and convolutional neural networks. Webology, 19(1): 1666-1688. <https://doi.org/10.14704/WEB/V19I1/WEB19112>

- [14] Monrose, F., Reiter, M.K., Li, Q., Wetzel, S. (2000). Cryptographic key generation from voice. In Proceedings 2001 IEEE Symposium on Security and Privacy. S&P 2001, Oakland, CA, USA, pp. 202-213. <https://doi.org/10.1109/SECPRI.2001.924299>
- [15] Khan, A.H., Aithal, P.S. (2022). Voice biometric systems for user identification and authentication—A literature review. *International Journal of Applied Engineering and Management Letters*, 6(1): 198-209. <https://doi.org/10.5281/zenodo.6471040>
- [16] Boles, A., Rad, P. (2017). Voice biometrics: Deep learning-based voiceprint authentication system. In 2017 12th system of systems engineering conference (SoSE), Waikoloa, HI, USA, pp. 1-6. <https://doi.org/10.1109/SYBOSE.2017.7994971>
- [17] Markowitz, J.A. (2000). Voice biometrics, who are you? Your voice along can be used to verify your personal identity-unobtrusively and invisibly. *Communications of the ACM*, 43(9): 66-73. <https://doi.org/10.1145/348941.348995>
- [18] Naji, S., Zainuddin, R., Kareem, S.A., Jalab, H.A. (2013). Detecting faces in colored images using multi-skin color models and neural network with texture analysis. *Malaysian Journal of Computer Science*, 26(2): 101-123.
- [19] Liu, Z., Huang, Z., Zhang, F. (2024). Improved AES algorithm based on dynamic row displacement and SHA-256 key extension. In *Fourth International Conference on Green Communication, Network, and Internet of Things (CNIoT 2024)*, pp. 55-63. <https://doi.org/10.1117/12.3052463>
- [20] Kaggle. Speaker Recognition Audio Dataset. <https://www.kaggle.com/datasets/vjcalling/speaker-recognition-audio-dataset/data>.
- [21] Stallings, W., Brown, L. (2023). *Computer Security: Principles and Practice*. Pearson.