Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7765732

MODIFIED PRIVACY FRAMEWORK BASED AUTOMATED VOICE AUTHENTICATION SYSTEM

Dr.M. Kathiresh[1], Dr.R. Sankarasubramanian[2]
[1]Assistant Professor, Department of Computer Applications, CMS College of Science and Commerce, Coimbatore, Tamil Nadu, India. Email: kathireshcs83@gmail.com
[2]Principal, Erode Arts and Science College, Erode. Email: rsankarprofessor@gmail.com

Abstract
There is always the risk that the audio recordings of the conversations will be sent through encrypted public networks and then kept and processed in the cloud by an external service provider. Voice-based interactive technologies are evolving towards an always-listening mode, inspired by the idea of continuous user involvement as opposed to being reliant on traditional push-to-talk capability. Rather than sharing the real voice model, this technological task is carried out through the construction of the chaff model that will be given to the server. Using the Gaussian mixture model, the chaff model of the voice signal is initially established. The created chaff model will be fed into the RSA model to generate the secret key pair. The resulting key pair will be used to encode the voice chaff model, which will then be shared with the server. The actual voice signal will be extracted at the server end using the appropriate keys, and it will then be authenticated using the convolutional neural network.
Keywords: Chaff Model, Gaussian Mixture Model, Neural Network, RSA and Voice Model

1. INTRODUCTION

A person communicates by repeating verbal, physiological, and auditory processes while speaking and listening. In recent years, speech recognition technology has advanced quickly, leveraging a variety of sensor technologies and bio signals that may be measured in these human voice activities. Particularly, the associated market has grown and is being integrated into a variety of services with the development of voice recognition technology based on artificial intelligence[1,2]. Speech recordings have abundance of personal, confidential data, which can be helpful in supporting different applications [3]. In spite of the privacy preservation, this has now been made mandatory by the current European and international data protection rules.

Speech signals include a wealth of personally identifiable information, most of which may be revealed using automated speaker and speech characterization methods [4]. Although both speaker and speech characterization are given attention in this study, the former is the primary emphasis[5]. The human voice belongs to the category of the most casual, non-invasive and easiest of all features [6,7]. In spite of the clear benefits and ubiquity of biometrics mechanism, doubts with respect to threats to privacy have crushed the trust of the public. Not just, the identity of a speaker is information that is highly confidential, the content spelled out might also be sensitive [8-10].

Privacy issues arise from the potential for interference and the inappropriate use of both biometric and non-biometric forms of speech data. Interceptions into personal privacy are

obviously non-permissible and the current EU General Data Protection Regulation is responsible for privacy preservation. Like mentioned earlier in this work, the persistent success enjoyed by speaker characterisation technology, and rather speech technologies on the whole, is found to rely on the advancement of robust and effective privacy preservation skills, particularly developed for the automated processing of speech signals[11-13].

2. PRIVACY CONCERNED VOICE AUTHENTICATION SYSTEM

Privacy based voice authentication is regarded to be the most important concept in the Defense sector to guarantee that voice communication and authentication is provided utmost security. In this article, it is accomplished through the generation of the chaff model that will be communicated to the server rather than sharing the actual voice model. In this, at first, the chaff model of voice signal is created with the help of the Gaussian mixture model. The chaff model created will be presented as input to the RSA model for the secret key pair generation. Depending on the created key pair, voice chaff model will be encoded and later shared with the server. In the server end, the actual voice signal will be obtained with the help of the right keys and thereafter, convolutional neural network will help in the authentication process. The characteristics for voice authentication will be obtained before using genetic based CNN that uses the Gabor filter. The entire processing flow of the newly introduced research approach is illustrated in the figure 1.

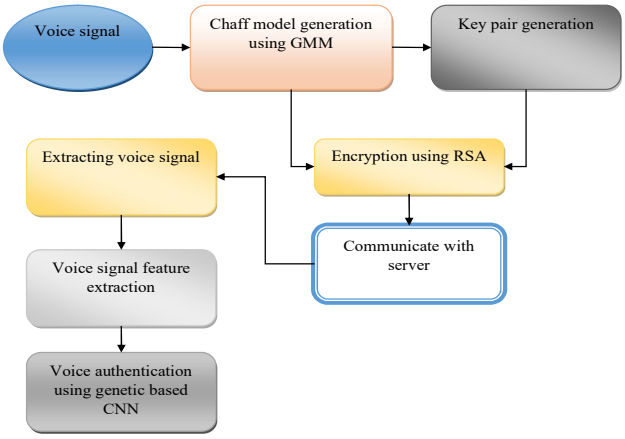


Figure 1. Processing flow of the proposed research work

3. USER REGISTRATION EMPLOYING RSA ENCRYPTION

To encrypt data, the RSA technique relies on the employment of a public key and a private key (two different but mathematically related keys). A private key should be kept secret and is never shown to anybody, while a public key is freely distributed. The RSA algorithm was named after its creators, Ron Rivest, Adi Shamir, and Leonard Adleman, who developed it in 1978. The below diagram shows the operation of asymmetric cryptography:

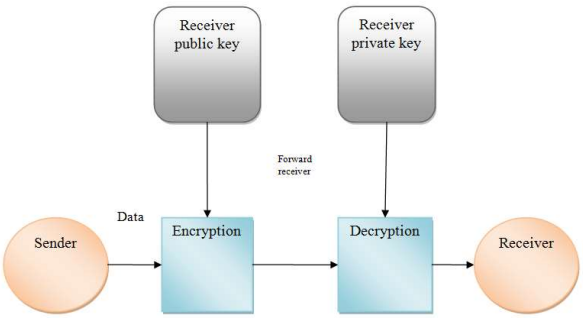


Figure 2 RSA processing flow

The RSA algorithm guarantees that the keys, given in the diagram above, are secure to the maximum possible.

3.1. User Enrolment and Verification Using RSA Encryption

The encryption key is completely different from (but related to) the decryption key. An integral part of the algorithm is modular exponentiation. The numbers e, d, and N are chosen to satisfy the condition that if A is a prime number between 1 and N, then (Ae mod N)d mod N = A. From this, we may infer that the user will employ e for encoding and d for decoding the letter A. On the flip side, the user will employ d for encoding and e for decoding (although performing this in a round manner is commonly referred as signing and verification). A public key is an openly accessible combination of the numbers (e,N). The combination (d,N) of numbers is the personal key and must be safeguarded. It is conventionally agreed that e is the public exponent, d is the private exponent, and N is the modulus. When compared to a symmetric block cypher like DES, RSA's performance degrades dramatically with increasing key length. As a result, the most straightforward option is to use RSA only for digital signatures and to encrypt DES keys using a different algorithm. The Data Encryption Standard (DES) must be used for mass data encryption.

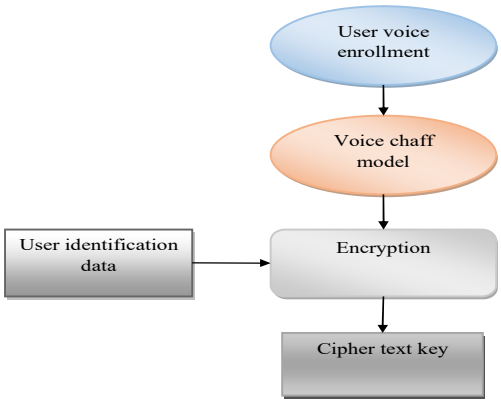


Figure 3. User enrolment and verification

4. VOICE AUTHENTICATION

The encrypted chaff voice model will be passed on to the server for carrying out the authentication. In the server end, decryption would be carried out and the voice authentication will be performed. In this research work, Convolution neural network is presented for the voice authentication step. In this, at first, the extraction of the voice features will be done employing the gabor filter that will later be learned applying the Convolutional neural network. The thorough explanations of these procedures are given in the sub sections that follow.

4.1. Convolutional Neural Network Based Voice Authentication

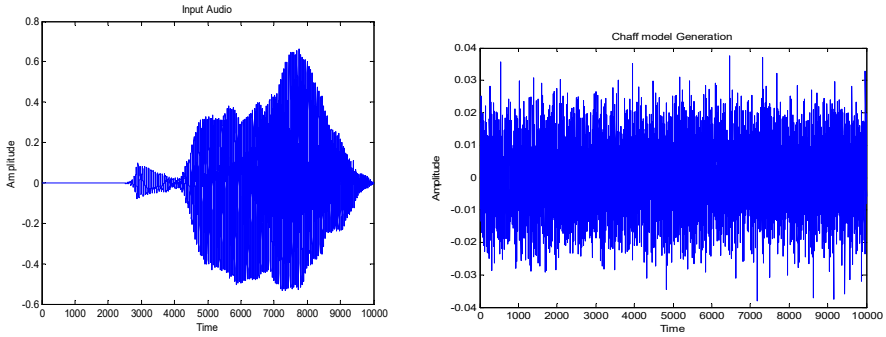
A convolutional neural network (CNN, or ConvNet) is a type of deep neural network typically employed in the study of visual images. Due to its shared-weights architecture and translation invariant properties, they are also known as shift invariant or space invariant artificial neural networks (SIANN). As a form of multilayer perception, CNNs are regularised versions. All neurons in one layer are connected to all neurons in the following layer, as is typical in such networks. For regularisation, CNNs also use a method that takes use of the hierarchical structure of data and combines more complex patterns with shorter, simpler ones. CNNs are towards the bottom in terms of connectivity and complexity.

5. RESULTS AND DISCUSSION

The discussed TLVAS assessment is carried out using the MAT Lab simulation environment. The voiceprint is then compared against each other for improving the performance.

5.1. Simulation Outcome

This section shows the simulation results achieved using the MAT Lab simulation environment. Figure shows the processing flow of the discussed research approach. In this technical work, at first, the chaff model of the input signal will be produced, whose encryption will be done with the secret key pair. Then this encrypted signal will be relayed to the received where it is decrypted and then feature extraction will be carried out. The features extracted will be learned employing the genetic based CNN algorithm for the authentication process.



(a) Input voice signal (b) Chaff model input signal

Figure 4 a,b shows the input voice signal used for the voice authentication process. This figure will be further processed to achieve voice authentication that is secure.

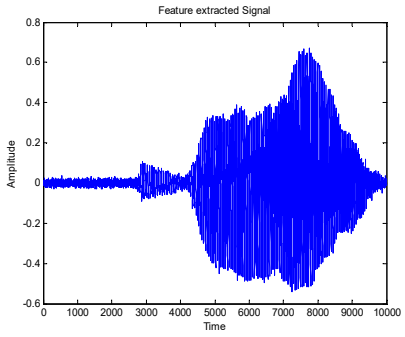


Figure 4(c) Feature extraction process

Figure 4c shows the feature extraction process carried out on the Chaff model of input signal. The learning of the extracted features will be done employing the genetic based CNN algorithm for the voice authentication results. This figure shows that the performance of the

proposed approaches can be quite good on voice authentication with improved level of accuracy.

5.2. Numerical Analysis

Sensitivity: Sensitivity is defined as the evaluation of ratio consisting of actual positive for categorizing the real voice to be the original. The sensitivity is defined as:

$$\text{Sensitivity} = \frac{T_p}{T_p + F_n} \tag{4}$$

Where $T_p$refers to the actual voice identified right to be the authenticated voice. $F_p$indicates the intruder voice, which is wrongly authenticated to be intruder voice. F_n, refers to the intruder voice is wrongly detected as the authenticated voice. The intruder voice T_n is identified correctly to be the intruder voice.

Precision: The accuracy of a voice signal can be calculated by comparing the number of correct identifications with the sum of the correct identifications and any false positives. The level of accuracy is defined as:

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{5}$$

Accuracy: Accuracy quantifies how well the model represents the world, and it is calculated as follows. the proportion of useful parameters for classification (T p+T n) to the full complement of parameters (T p+T n+F p+F n).

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{6}$$

Signal to Noise Ratio: The signal-to-noise ratio is a metric used in the sciences and engineering to evaluate the strength of a signal relative to the strength of any accompanying noise. SNR is often measured in dB and is defined as the ratio of signal power to noise power. There is more signal than noise if the signal-to-noise ratio is larger than 1.

$$\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}} \tag{7}$$

Root Mean Square Deviation: For the comparison of the encryption results with less fluctuation in the parameter, the root mean square difference (RMSD) is calculated. Let f μ1,x1(0) refer to the encryption result of the signal and let f μ2,x2(0) be the non encrypted signal; the RMSD between f μ1,x1(0) and f μ2,x2(0) is expressed as

$$\text{RMSD} = \left(\frac{1}{L \times P} \sum_{i=0}^{L-1} \sum_{j=0}^{P-1} \left(f'_{\mu_1 x_1(0)}(i,j) - f'_{\mu_2 x_2(0)}(i,j) \right)^2 \right)^{1/2} \tag{8}$$

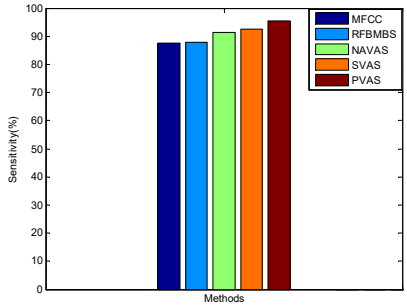
Non-invertibility Index (NI): With the aim ofsafeguarding the confidential data, few algorithms might change biometric information. These changes has to be unchangeable, so that it can be guaranteed that if there is an attack on a biometric storage repository, the intruders cannot retrieve the actual personal biometric information of the user using the data maintained in the database

Table 1 Performance metric comparison values

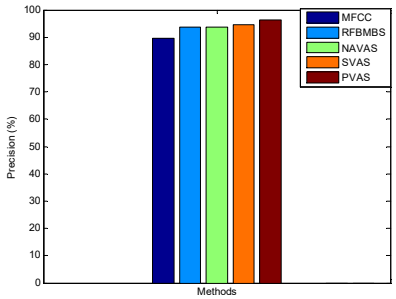
Methods	Sensitivity (%)	Precision (%)	Accuracy (%)	SNR (dB)	RMSD
MFCC	87.50	89.74	82.00	7.78	0.94

RFBMBS	87.90	93.59	85.00	9.13	0.87
NAVAS	91.50	93.83	88.00	13.33	0.71
SVAS	92.40	94.60	91.00	14.78	0.65
PVAS	95.60	96.20	97.50	16.70	0.43

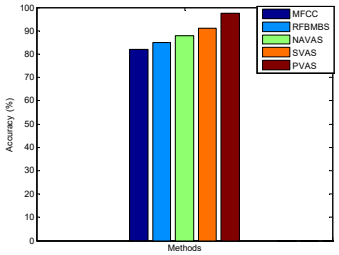
After performing a sensitivity analysis, it becomes clear that the recently presented method outperforms the existing ones. The results of evaluation are given as below: for MFCC it is 87.5%, Retina and Fingerprint Based Multi –Biometric System (RFBMBS) technique yields 87.9%, NAVAS achieves 91.5%, SVAS yields 92.4% and PVAS achieves 95.6%. The performance newly introduced techniques in terms of this precision assessment outperforms theavailable techniques. The MFCC yields 89.7%,RFBMBSyields 93.5% , NAVAS renders93.8%. 94.6% achieves SVAS and PVAS achieves 96.2%. The evaluation of the performance of the discussed technique in terms of accuracy is much better than the available techniques. MFCC technique yields 82.00%, 85.00% yields RFBMBS, NAVAS gives 88%,SVAS gives 91% and PVAS renders 97.5%. When evaluated in terms of SNR, the discussed technique PVAS achieves superior performance whereas it is 12.99% more than SVAS, 25.28% more than NAVAS, 82.91% more than RFBMBS and 114.65% more than MFCC. In terms of RMSD metric, PVAS demonstrates superior performance where it is 33.84% lesser compared to SVAS, 39.43% compared to NAVAS, 50.57% compared to EFBMBS and 54.25% compared to MFCC.



5a. Performance Evaluation of Accuracy with Different Methodswith Different Methods



5b.Performance Evaluation of Accuracy



5c. Performance Evaluation of Accuracy with Different Methods

6.CONCLUSION

The Privacy related voice authentication. In this technical work, it is carried out through the creation of the chaff model that will be passed to the server rather than having the actual

voice model shared. In this, at first, chaff model of voice signal is established using the Gaussian mixture model. The chaff model generated will be provided as input to the RSA model for creating the secret key pair. As per the generated key pair, voice chaff model will be encoded and thereafter shared with the server. In the server end, the extraction of the actual voice signal will be done utilizing the right keys and after this, it will be authenticated employing the convolutional neural network. The features used for voice authentication will be obtained prior to the application of genetic based CNN that uses the Gabor filter.

REFERENCES

[1] Ammari, T., Kaye, J., Tsai, J. Y., & Bentley, F. (2019). Music, search, and IoT: How people (really) use voice assistants. ACM Transactions on Computer-Human Interaction (TOCHI), 26(3), 1-28.

[2] Nautsch, A., Jiménez, A., Treiber, A., Kolberg, J., Jasserand, C., Kindt, E., Abdelraheem, M. A. (2019). Preserving privacy in speaker and speech characterisation. Computer Speech & Language, 58, 441-480.

[3] Rui, Z., & Yan, Z. (2018). A survey on biometric authentication: Toward secure and privacy-preserving identification. IEEE Access, 7, 5994-6009.Language Processing, Vol.20, No.8, Pp.2280-2290, 2012.

[4] Liaqat, D., Nemati, E., Rahman, M., &Kuang, J. (2017, December). A method for preserving privacy during audio recordings by filtering speech. In 2017 IEEE Life Sciences Conference (LSC) (pp. 79-82). IEEE.

[5] Jahan, S., Chowdhury, M., & Islam, R. (2019). Robust user authentication model for securing electronic healthcare system using fingerprint biometrics. International Journal of Computers and Applications, 41(3), 233-242.

[6] Xing, Y., Wang, T., Zhou, F., Hu, A., Li, G., & Peng, L. (2020). EVAL cane: Non-intrusive monitoring platform with a novel gait-based user identification scheme. IEEE Transactions on Instrumentation and Measurement.

[7] Kyriakopoulos, K., Gales, M., &Knill, K. (2018). Automatic characterisation of the pronunciation of non-native English speakers using phone distance features.

[8] Yoshioka, T., Erdogan, H., Chen, Z., &Alleva, F. (2018, April). Multi-microphone neural speech separation for far-field multi-talker speech recognition. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5739-5743). IEEE.

[9] Joshi, M., Mazumdar, B., & Dey, S. (2020). A Comprehensive Security Analysis of Match-in-database Fingerprint Biometric System. Pattern Recognition Letters.

[10] Van der Schyff, K., Flowerday, S., & Furnell, S. (2020). Duplicitous Social Media and Data Surveillance: An evaluation of privacy risk. Computers & Security, 101822.

[11] Deng, J., Guo, J., Zhang, D., Deng, Y., Lu, X., & Shi, S. (2019). Lightweight face recognition challenge. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 0-0).

[12] Nguyen, K., Fookes, C., Ross, A., & Sridharan, S. (2017). Iris recognition with off-the-shelf CNN features: A deep learning perspective. IEEE Access, 6, 18848-18855.

[13] Jain, A. K., Arora, S. S., Cao, K., Best-Rowden, L., & Bhatnagar, A. (2016). Fingerprint recognition of young children. IEEE Transactions on Information Forensics and Security, 12(7), 1501-1514.