

Authentication of air traffic controllers in LDACS networks based on voice biometrics

Georgiy Konakhovych^{1,†}, Oleksii Holubnychyi^{1,†} and Semen Doroshenko^{2,*,†}

¹Scientific Cyber Security Association of Ukraine, Mykhaila Donsia Str., 2A, Kyiv, 03161, Ukraine

²LLC Cast Iron MT, Ioanna Pavla II Str., 7, Kyiv, 01042, Ukraine

Abstract

The current global security situation requires additional measures to gain the highest aviation safety. Here proposed an authentication method for air traffic controllers in the perspective LDACS aeronautical communication network by using a voice characteristic. The main idea is to combine ECAPA-TDNN speaker feature extraction, binary template generation, the Fuzzy Commitment scheme, and error-correcting codes. The system complements conventional ATN/IPS security mechanisms by verifying the physical identity of the controller rather than only the hardware network or terminating device.

Keywords

cryptobiometric authentication; LDACS; ATN/IPS; machine learning; deep learning; neural networks; ECAPA-TDNN; speaker verification; Fuzzy Commitment; voice biometrics; aviation cybersecurity

1. Introduction

The world nowadays is full of unpredictable military and terrorist threats, especially in Ukraine. Even the war stops the real threats will be kept at the highest level for a long time. Therefore, the cyber security issues are the most important wherever they are considered. At the same time modern aeronautical telecommunication infrastructure is undergoing a transition from analogue voice communication systems to digital data transmission systems based on the Internet Protocol (IP). This transition is driven by the growth of air traffic intensity, increased requirements for channel capacity, the need to support automated data exchange between aircraft and ground services, and the integration of air traffic management systems into a unified digital environment.

One of the key directions of this modernization is the deployment of the Aeronautical Telecommunication Network using Internet Protocol Suite (ATN/IPS), which provides the use of IP-based protocols in safety-critical aeronautical services. Within this concept, LDACS (L-band Digital Aeronautical Communications System) is considered a promising digital air-ground communication system. LDACS is a broadband, spectrally efficient and reliable aeronautical communication system operating in the L-band and supporting IPv6-based data transmission [1, 2, 3, 4, 5, 6].

Unlike conventional analogue voice radio communication, LDACS has a scheduled medium access structure, supports digital services, service messages, aeronautical operational data, and potentially digital voice. At the same time, the transition to open IP-oriented architectures introduces new cybersecurity risks [7]. In analogue systems, the main threats were interference, unauthorized listening, and voice imitation. In IP-oriented networks, additional threats include spoofing, replay attacks, man-in-the-middle attacks, key compromise, certificate substitution, and unauthorized access to service channels.

Standard security mechanisms for ATN/IPS and LDACS rely on cryptographic authentication of

CMiGIN'26: Fifth International Workshop on Conflict Management in Global Information Networks, June 26–27, 2026, Krakow, Poland

*Corresponding author.

[†]These authors contributed equally.

✉ heorhii.konakhovych@npp.kai.edu.ua (G. Konakhovych); oleksii.holubnychyi@npp.kai.edu.ua (O. Holubnychyi); semendoroshenko@gmail.com (S. Doroshenko)

ORCID 0000-0002-6636-542X (G. Konakhovych); 0000-0001-5101-3862 (O. Holubnychyi); 0009-0007-1302-3884 (S. Doroshenko)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

network nodes, public key infrastructure (PKI), X.509 digital certificates, IPsec/IKEv2 protocols, and secure channel establishment procedures [8, 9, 10, 11]. These mechanisms are important and necessary as they primarily authenticate a terminating device or network node, however, not the physical person initiating a transmission. If equipment, credentials, or cryptographic certificates are compromised, the system may not be able to distinguish a legitimate air traffic controller from an attacking person who has gained access to an authorized terminal [12, 13, 14].

In this context, it is relevant to introduce an additional authentication layer focused on the operator as a physical person. One such approach is using unique voice characteristics. A related direction involving speech signal analysis in an aviation security procedure was previously considered in a Ukrainian patent [15]. The controller's voice is a natural and continuously available biometric factor during radio communication. Unlike passwords, tokens, or certificates, voice biometrics are directly linked to a physical person. However, the use of voice data in aeronautical communication is complicated by the high variability of speech signals caused by noise, stress, communication channel conditions, headset type, emotional state, speech rate, and other factors.

Cryptographic algorithms, in contrast, require exact reproduction of keys. Even a single-bit change causes a complete change in a hash value or encryption result. This creates a fundamental contradiction: biometric data are naturally fuzzy, while cryptographic keys must be deterministic. To overcome this contradiction, crypto-biometric systems are used, including the Fuzzy Commitment Scheme and Fuzzy Extractors, which make it possible to recover a stable cryptographic key from noisy biometric data [16, 17, 18, 19].

The purpose of this work is to develop and provide a theoretical and experimental justification for a crypto-biometric authentication method for air traffic controllers in LDACS networks based on neural-network extraction of voice features using ECAPA-TDNN, the Fuzzy Commitment Scheme, and error-correcting coding.

The object of the study is the process of authenticating an air traffic controller in a digital aeronautical telecommunication network.

The subject of the study is the methods for generating, protecting, and recovering a cryptographic key based on voice biometric features in a noisy aeronautical communication channel.

The scientific novelty of the work lies in combining neural-network speaker embeddings, the crypto-biometric Fuzzy Commitment Scheme, and the architectural integration of this mechanism into the LDACS digital aeronautical network as an additional layer of personal authentication for air traffic controllers in the context of ATN/IPS modernization.

2. Analysis of recent research and publications

2.1. Digital aeronautical communication: LDACS and ATN/IPS

LDACS is a promising digital aeronautical communication system operating in the L-band and intended to provide a reliable air-ground link. It is described as a secure, scalable, and spectrally efficient data transmission system for civil aviation with IPv6 support [1]. Technical LDACS specifications provide for OFDM/OFDMA-based approaches, medium access, control service channels, and support for aeronautical services within the requirements of ATN/IPS [5, 6, 9, 20].

The physical layer of LDACS uses orthogonal frequency division multiplexing, which improves spectral efficiency and resistance to multipath propagation. In general form, an OFDM signal can be represented as:

$$s(t) = \sum_{k=0}^{N-1} X_k e^{j2\pi k \Delta f t}, 0 \leq t < T_s,$$

where X_k is the complex symbol on the k -th subcarrier, N is the number of subcarriers, Δf is the frequency spacing between subcarriers, and T_s is the OFDM symbol duration.

The orthogonality of subcarriers is ensured by the condition:

$$\Delta f = \frac{1}{T_s}$$

For the aeronautical channel, the Doppler shift caused by high aircraft speed is important:

$$f_D = \frac{v}{c} f_c \cos \theta,$$

where v is the relative aircraft velocity, c is the speed of light, f_c is the carrier frequency, and θ is the angle between the direction of motion and the direction of wave propagation.

Therefore, an authentication algorithm integrated into LDACS must be also computationally lightweight so that it does not introduce excessive delay into service channels [8].

2.2. Voice biometrics and speaker verification

Automatic Speaker Verification (ASV) verifies a speaker's identity based on their voice. Unlike speech recognition, where the crucial objective is to determine the content of an utterance, speaker verification analyzes the individual characteristics of the vocal tract and personal speaking style.

Well-known speaker verification systems were based on GMM-UBM, i-vector, and PLDA approaches. Modern systems are dominated by neural-network embedding models, in particular x-vector and ECAPA-TDNN [21, 22, 23]. General principles of biometric recognition, including identification and verification tasks, are described in [24].

The general task of speaker verification can be formulated as follows. Let x be a test audio signal, and let E_{ref} be the reference feature vector of the claimed speaker. A neural-network extractor transforms the audio signal into a feature vector:

$$E = F_\theta(x), E \in \mathbb{R}^d.$$

For ECAPA-TDNN, a typical embedding dimension is:

$$d = 192$$

Similarity between the reference and test embeddings is usually estimated by cosine similarity:

$$S_C(E_1, E_2) = \frac{E_1 \cdot E_2}{\|E_1\| \|E_2\|}.$$

The acceptance or rejection decision is made according to the rule:

$$\begin{aligned} &\text{Accept, if } S_C(E_{ref}, E_{test}) \geq \tau, \\ &\text{Reject, if } S_C(E_{ref}, E_{test}) < \tau, \end{aligned}$$

where τ is the decision threshold.

ECAPA-TDNN extends the TDNN and x-vector concepts through several mechanisms:

- Res2Net-like blocks for multi-scale analysis of temporal dependencies;
- Squeeze-and-Excitation modules for modeling inter-channel dependencies;
- attentive statistics pooling for weighted aggregation of temporal frames;
- aggregation of features from different levels of the neural-network hierarchy [22].

In general form, attentive statistics pooling can be described as:

$$\begin{aligned} \mu &= \sum_{t=1}^T \alpha_t h_t, \\ \sigma &= \sqrt{\sum_{t=1}^T \alpha_t h_t^2 - \mu^2}, \end{aligned}$$

where h_t is the hidden feature vector at time t , α_t is the attention weight, μ is the weighted mean, σ is the weighted standard deviation.

2.3. Protection of biometric templates

Biometric systems differ fundamentally from password systems because biometric characteristics cannot be reissued in case of compromising. A breached password can be changed, whereas a compromised voice or other biometric template cannot be fully replaced. This is a core problem of biometric systems because biometric characteristics are permanently associated with a physical person [24].

ISO/IEC 24745 defines key requirements for biometric information protection, including [25]:

- confidentiality;
- integrity;
- irreversibility of the original biometric template;
- renewability or revocability of the protected representation;
- unlinkability between different applications of the same biometric source.

Storing embedding vectors in open form is unsafe because they can be used for unauthorized identification, attack construction, or partial reconstruction of biometric properties. Therefore, a biometric template should be stored only in protected, irreversible, or cryptographically bound form. The performance testing of biometric template protection schemes is addressed in [26].

2.4. Fuzzy commitment and fuzzy extractor

The Fuzzy Commitment Scheme was proposed in [16] as a cryptographic scheme that allows a secret to be bound to a noisy biometric template. This idea was further developed in [17, 18], where the concept of a fuzzy extractor was formalized.

The classical fuzzy extractor model consists of procedures:

$$\begin{aligned} Gen(w) &\rightarrow (R, P) \\ Rep(w', P) &\rightarrow R, \end{aligned}$$

where w is the initial biometric template, w' is a new template of the same person, R is the recovered cryptographic key, and P is helper data.

The correctness condition is:

$$\text{if } dist(w, w') \leq t, \text{ then } Rep(w', P) = R$$

The security condition requires that helper data should not reveal either the biometric template or the cryptographic key:

$$P \Rightarrow \neg w, P \Rightarrow \neg R.$$

In the Fuzzy Commitment Scheme, a secret key or codeword is masked by a biometric template:

$$P = C \oplus w,$$

where C is the codeword of an error-correcting code, w is the binary biometric template, and P is helper data.

During authentication, a new template w' is formed, and then:

$$C' = P \oplus w' = C \oplus w \oplus w' = C \oplus e,$$

where e is the biometric noise vector:

$$e = w \oplus w'$$

If the error-correcting code can correct the noise e , the original codeword C and therefore the secret key can be recovered.

2.5. Reed-Solomon codes and limitations of their use

Reed-Solomon codes are block error-correcting codes over finite Galois fields [27, 28]. A classical code is specified by the parameters $RS(n, k)$, where n is the codeword length, k is the information message length, and $n-k$ is the number of redundant symbols.

The maximum number of correctable symbol errors is:

$$t = \left\lfloor \frac{n - k}{2} \right\rfloor.$$

It is important to emphasize that a Reed-Solomon code operates on symbols rather than individual bits. In practical implementations, one symbol often corresponds to one byte. Therefore, if biometric noise is distributed across many bytes, even a relatively small number of bit errors may cause a large number of symbol errors.

This is a fundamental limitation when applying Reed-Solomon codes in biometric systems. For voice embeddings, where post-binarization errors often have a diffuse character, it is also appropriate to consider alternatives correction algorithms such as:

- BCH codes;
- LDPC codes;
- Polar codes;
- secure sketches based on Hamming distance;
- cascade error-correction schemes.

Under this research Reed-Solomon coding is considered as a baseline simulation mechanism for error correction that allows the principle feasibility of key recovery from an unstable voice template to be evaluated. The final choice of the code for practical aeronautical implementation requires separate optimization.

3. Mathematical model of proposed method

3.1. Speech signal model

The speech signal of an air traffic controller can be represented as the result of interaction between an excitation source and the vocal tract. This model corresponds to the classical approach to speech analysis in speaker recognition tasks [21, 22, 24]:

$$s(t) = e(t) * h(t) + n(t),$$

where $g(t)$ is the excitation signal of the vocal cords, $h(t)$ is the impulse response of the vocal tract, $n(t)$ is the noise component, and $*$ denotes convolution.

In the frequency domain:

$$S(f) = E(f)H(f) + N(f).$$

For aeronautical radio communication, the noise component may include several components, such as background noise in the controller workplace, headset noise, quantization noise, bandwidth limitations, packet losses, and channel distortions [1, 2, 3, 4, 5, 6, 8, 29, 30].

The signal-to-noise ratio is defined as:

$$SNR_{dB} = 10 \log_{10} \frac{P_s}{P_n},$$

where P_s is the power of the useful signal and P_n is the noise power.

3.2. Feature vector formation

After preprocessing, the audio signal is transformed into a time-frequency representation using mel spectrograms [21, 22]:

$$M = Mel(|STFT(x)|^2),$$

where $STFT$ is the short-time Fourier transform and Mel denotes transformation to the mel scale. The neural-network extractor forms a speaker feature vector:

$$e_i = F_\theta(x_i), e_i \in \mathbb{R}^{192}.$$

In the practical implementation, this stage can be performed using a pretrained ECAPA-TDNN model implemented in SpeechBrain [31, 32].

To reduce random variability, a centroid of enrollment utterances is formed:

$$e_{enr} = \frac{1}{N} \sum_{i=1}^N e_i$$

The centroid approach reduces the variance of random deviations:

$$Var(e_{enr}) = \frac{\sigma^2}{N},$$

provided that the noise components are statistically independent.

3.3. Binarization of biometric features

To use a real-valued embedding vector in the Fuzzy Commitment Scheme, it must be transformed into a binary sequence. This is necessary because the classical Fuzzy Commitment Scheme works with discrete binary templates and the XOR operation [16, 17, 18, 19]:

$$B : \mathbb{R}^d \rightarrow \{0, 1\}^d.$$

A basic binarization rule can be written as:

$$w_i = \begin{cases} 1, & e_i > \mu_i, \\ 0, & e_i \leq \mu_i, \end{cases}$$

where μ_i is the threshold for the i -th feature.

The threshold may be defined as the median of the current embedding, a global mean estimated from the training set, a feature-wise statistical threshold, or a parameter optimized according to FAR/FRR. The quality of such transformations should be evaluated with regard to biometric template protection requirements and false acceptance/false rejection metrics [25, 26].

To improve stability, normalization is applied:

$$\hat{e}_i = \frac{e_i - \mu_i}{\sigma_i},$$

and then:

$$w_i = \mathbb{I}(\hat{e}_i > 0).$$

The distance between two binary templates is evaluated using Hamming distance [25, 26]:

$$d_H(w, w') = \sum_{i=1}^d w_i \oplus w'_i.$$

The normalized Hamming distance is:

$$D_H(w, w') = \frac{d_H(w, w')}{d}$$

If the distance for the same speaker is significantly lower than the distance for different speakers, the system has potential for reliable biometric authentication.

3.4. Formation of the fuzzy commitment

At the enrollment stage, a random cryptographic key is generated [16, 17, 18, 19]:

$$K \leftarrow \{0, 1\}^\uparrow$$

The key is then encoded by an error-correcting code, in particular by Reed-Solomon codes [27, 28]:

$$C = Enc(K).$$

After that, helper data are formed:

$$P = C \oplus w.$$

The system memory stores:

$$DB = \{P, H(K), ID\},$$

where P is helper data, $H(K)$ is the cryptographic hash of the key, and ID is the user identifier. The biometric template and the key are not stored in open form, which is consistent with biometric template protection principles [19, 25, 26].

3.5. Key recovery during authentication

During authentication, a new embedding e' and then a new binary template w' are formed. Next:

$$C' = P \oplus w'.$$

The resulting C' is a noisy codeword that can be recovered by an error-correcting decoder if the number of errors does not exceed the correction capability of the code [16, 27, 28].

Since:

$$P = C \oplus w,$$

then:

$$C' = C \oplus w \oplus w' = C \oplus e,$$

where e is the biometric noise vector.

If the condition

$$d_H(w, w') \leq t$$

is satisfied, the decoder recovers:

$$K' = Dec(C') = K.$$

Then a hash verification is performed:

$$H(K') = H(K).$$

If the equality holds, authentication is successful. Storing only the key hash and helper data reduces the risk of biometric information compromise and is consistent with the principles of biometric template protection [19, 25, 26].

4. Integration into the LDACS architecture

The proposed technology does not replace standard PKI, IPsec and IKEv2 mechanisms. Rather, it supplements them with a personal biometric factor. It should be considered an additional security layer above the basic network infrastructure. This approach is consistent with the general concept of phased LDACS deployment as a next-generation digital aeronautical communication system [5, 6].

A possible operational sequence is as follows:

1. The air traffic controller initiates a service communication session.
2. The system records a short speech fragment.
3. ECAPA-TDNN forms a speaker embedding.
4. Binarization and key recovery are performed.
5. If verification is successful, the key is used as an additional factor for initiating a protected session.
6. The service message is transmitted through the LDACS channel.

The size of helper data P depends on the length of the binary biometric template and on the parameters of error-correcting coding. In the proposed scheme, ECAPA-TDNN forms a feature vector of dimension:

$$e \in R^{192}$$

If each component of the feature vector corresponds to one bit after binarization, a binary template is formed:

$$w \in \{0, 1\}^{192}$$

In this baseline case, its size is:

$$192 \text{ bits} = 192/8 = 24 \text{ bytes}.$$

The actual size of helper data may be larger if error-correcting coding or projection-based binarization with a larger number of bits is applied. Even with redundancy, the amount of service data remains relatively small compared with typical network packets. This allows such a mechanism to be integrated into service messages without substantial channel overload.

Since aeronautical data communication services have strict requirements for safety, reliability, and message delivery time, the additional authentication mechanism must not introduce excessive delay [8]. The total authentication delay can be expressed as:

$$T_{auth} = T_{rec} + T_{prep} + T_{emb} + T_{bin} + T_{dec} + T_{net},$$

where T_{rec} is the time for recording the speech fragment, T_{prep} is preprocessing time, T_{emb} is embedding extraction time, T_{bin} is binarization time, T_{dec} is decoding time, and T_{net} is network delay.

For practical use, the following condition should hold:

$$T_{auth} \leq T_{max}$$

where T_{max} is the maximum permissible delay for the corresponding service scenario.

5. Experimental methodology

5.1. Input data

The experimental implementation of the proposed method was performed in Python. The verification used the ATCOSIM resource, which contains air traffic controller speech in WAV format under simulated air traffic control conditions [29, 30]. This corpus is relevant because aeronautical radio communication has specific vocabulary, phraseology, speech tempo, and acoustic characteristics. A dataset was prepared for seven speakers (D1-D7). For each speaker, 10 recordings were used for enrollment and 30 independent recordings were used for testing. Each test recording was compared with the reference template of each speaker, which resulted in $7 \times 30 \times 7 = 1470$ pairwise checks.

5.2. Preprocessing

Each audio file passed through the following stages:

1. Conversion to mono format, because the aviation voice channel is single-channel and the neural network processes monaural audio.
2. Simulation of a narrowband voice channel as the audio signal was converted to 8 kHz sampling frequency with 8-bit quantization, which corresponds to a bitrate of 64 kbit/s.
3. Resampling from 8 kHz to 16 kHz only to ensure compatibility with the pretrained ECAPA-TDNN model. This resampling does not restore spectral components lost due to the 8 kHz channel limitation.
4. Amplitude normalization.

$$x_{norm}(t) = \frac{x(t)}{\max |x(t)|}$$

5. Silence removal using an energy threshold.
6. Formation of a mel spectrogram.
7. Extraction of the feature vector using a pretrained ECAPA-TDNN model implemented in Speech-Brain [31, 32].

5.3. Testing scenarios

The following scenarios were used in the experimental study:

- Scenario A — legitimate speaker. New recordings of the same speaker that were not used in the enrollment stage are used.
- Scenario B — impostor speaker. Recordings of other operators are used.

5.4. Evaluation metrics

The main metrics are False Acceptance Rate (FAR), False Rejection Rate (FRR), and Equal Error Rate (EER).

$$FAR = \frac{N_{FA}}{N_{imp}} \cdot 100\%,$$

where N_{FA} is the number of falsely accepted impostor attempts and N_{imp} is the total number of impostor attempts.

$$FRR = \frac{N_{FR}}{N_{leg}} \cdot 100\%,$$

where N_{FR} is the number of falsely rejected legitimate attempts and N_{leg} is the total number of legitimate attempts.

$$FAR(\tau) = FRR(\tau)$$

The Bit Error Rate after binarization is also evaluated:

$$BER = \frac{d_H(w, w')}{d}$$

For the crypto-biometric module, the following characteristics should be evaluated:

- successful key recovery rate;
- average key recovery time;
- number of decoding errors;
- resistance to impostor speakers;

6. Results and interpretation

6.1. Analysis of cosine similarity between controllers

For experimental verification of the proposed system, a set of voice samples from several air traffic controllers denoted as D1-D7 was formed. For each controller, a reference embedding vector was created, after which test samples were compared pairwise with the reference templates of all controllers. Evaluation was carried out in two related dimensions: first, cosine similarity between ECAPA-TDNN embedding vectors was analyzed; second, Hamming distance between binarized templates and the ability of the crypto-biometric module to recover the key were investigated.

At the first stage, a matrix of mean cosine similarity between test and reference voice samples of controllers was constructed. The results are shown in Fig. 1.

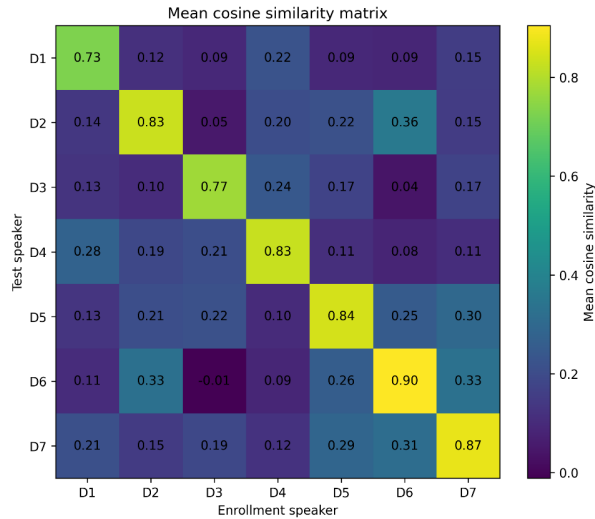


Figure 1: Cosine similarity matrix between controllers

The matrix shows a clear concentration of high values on the main diagonal, corresponding to comparisons of samples from the same controller. For legitimate comparisons, cosine similarity values are mostly within the range 0.73-0.90.

The highest mean value was obtained for controller D6, $S_C = 0.90$ which indicates high stability of this controller's voice embedding vectors within the experimental sample. Other controllers also demonstrate high diagonal values: D2 = 0.83, D4 = 0.83, D5 = 0.84, and D7 = 0.87.

Outside the main diagonal, the cosine similarity values are significantly lower. Most inter-speaker comparisons fall approximately within the range 0.01-0.36. This indicates sufficient discriminative

ability of ECAPA-TDNN embeddings for separating a legitimate controller from impostor speakers. The highest inter-speaker values are observed for the pairs D2-D6, D6-D7, and D5-D7, which may indicate partial acoustic or timbral similarity of individual voices. However, these values remain substantially lower than the diagonal values, allowing effective separation of genuine and impostor comparisons.

6.2. Distribution of genuine and impostor scores

For a more detailed analysis, two comparison classes were formed: genuine scores, which correspond to comparisons of samples from the same controller, and impostor scores, which correspond to comparisons of samples from different controllers. The distributions are shown in Fig. 2.

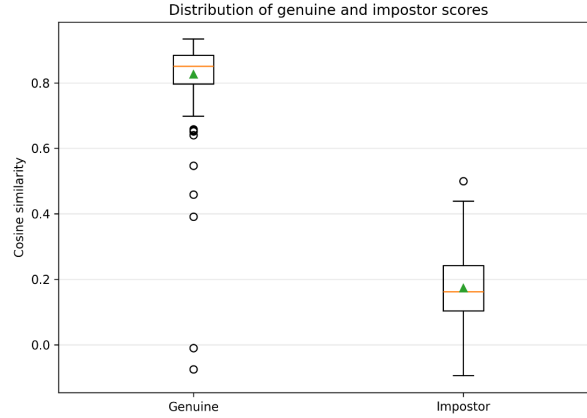


Figure 2: Distribution of genuine and impostor scores

The boxplot shows that genuine scores are substantially higher than impostor scores. The median of the genuine distribution is approximately 0.84, whereas the median of the impostor distribution is approximately 0.17. Thus, a pronounced separation interval is formed between the two classes.

This conclusion is also confirmed by the histogram of cosine similarity distributions shown in Fig. 3.

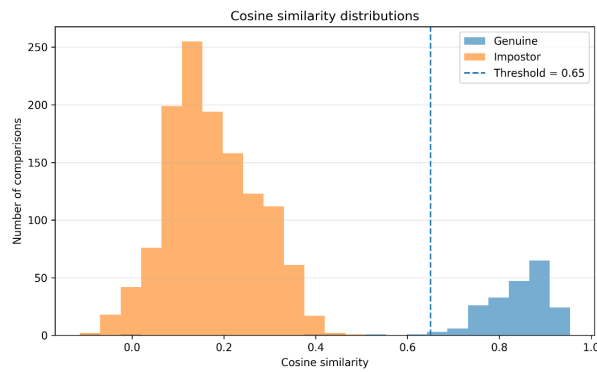


Figure 3: Histogram of cosine similarity distributions

Genuine values are mainly concentrated in the range from 0.75 to 0.95, whereas impostor values are mostly below 0.4. The histogram also marks the operational threshold $\tau = 0.65$. With this threshold, most legitimate comparisons are accepted by the system, while impostor speakers are blocked. This makes it possible to use 0.65 as a practical threshold in the demonstration configuration. At the same time, further FAR/FRR analysis shows that the statistically optimal threshold may be lower.

6.3. ROC analysis of the voice authentication system

For quantitative evaluation of classification quality, an ROC curve was constructed. The ROC curve shows the relationship between True Positive Rate and False Positive Rate when the decision threshold changes. The ROC curve is shown in Fig. 4.

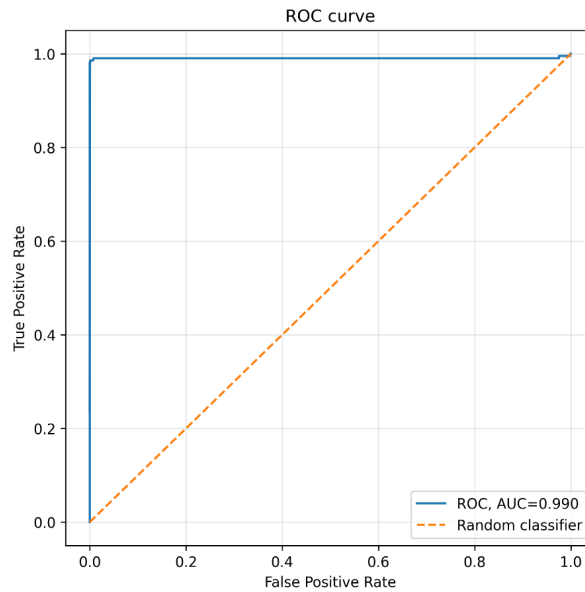


Figure 4: ROC curve

The area under the ROC curve is $AUC = 0.99$

This value indicates very high discriminative ability. The curve is located close to the upper-left corner, meaning that the system is capable of achieving a high rate of correct acceptance of legitimate users with a low rate of false acceptance of impostor speakers. Compared with a random classifier, which corresponds to the diagonal of the ROC space, the proposed system demonstrates substantially better class separation. This confirms the effectiveness of ECAPA-TDNN embeddings for controller authentication.

6.4. FAR and FRR analysis depending on the threshold

To determine the optimal threshold, FAR and FRR dependencies on the decision threshold were constructed. The results are shown in Fig. 5.

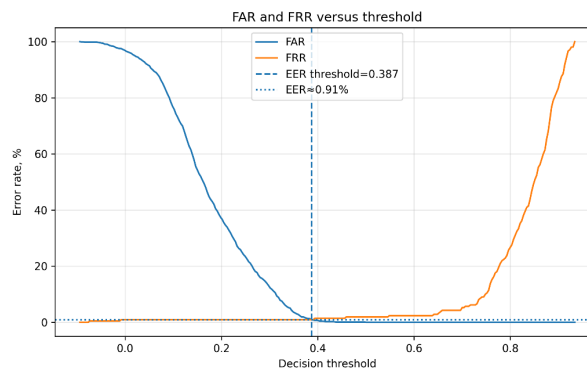


Figure 5: FAR and FRR versus decision threshold

As the threshold increases, FAR gradually decreases because the system becomes stricter in accepting

users. At the same time, FRR increases because some legitimate users begin to be falsely rejected. The point where FAR equals FRR corresponds to the Equal Error Rate. The experimental results show:

$$EER \approx 0.91\%.$$

The corresponding threshold is:

$$\tau_{EER} \approx 0.387$$

The low EER value indicates high biometric class separation quality. However, in a practical aeronautical system, the threshold does not necessarily have to be equal to the EER threshold. In safety-critical systems, it is advisable to use a more conservative threshold that reduces the probability of false acceptance of an impostor, even if this slightly increases the probability of false rejection of a legitimate controller. Therefore, two thresholds may be considered in the demonstration configuration:

$$\tau_{EER} \approx 0.387$$

as a statistically optimal threshold and $\tau = 0.65$ as a stricter operational threshold for increased security.

6.5. Analysis of Hamming distance between binarized templates

The crypto-biometric part of the system was studied separately. For this purpose, embedding vectors were converted into binary templates, and the normalized Hamming distance between templates of different controllers was calculated. The results are shown in Fig. 6.

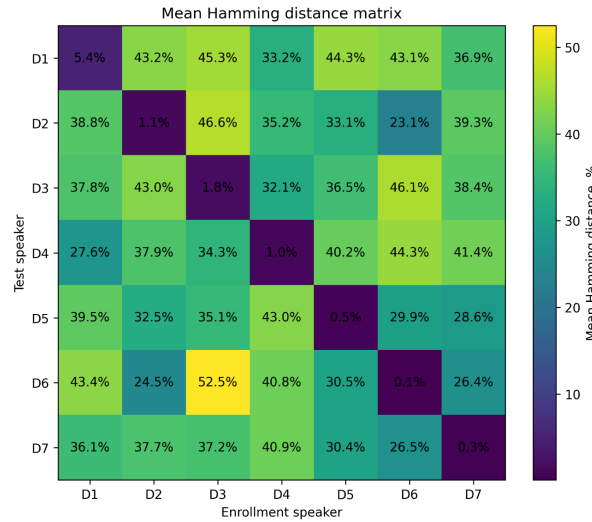


Figure 6: Hamming distance matrix between binarized template

Low values are observed on the main diagonal of the Hamming distance matrix: approximately 0.1-5.4%. This corresponds to comparisons of templates of the same controller and indicates sufficient stability of the binarized representation after feature stabilization.

Outside the main diagonal, the Hamming distance values are significantly higher. For inter-speaker comparisons, they mainly fall within the range 23.1-52.5%. The largest inter-speaker distance is observed for the pair D6-D3: $D_H = 52.5\%$

This is close to the expected level of discrepancy between independent random binary templates, for which the mean Hamming distance approaches 50%. Thus, the obtained results confirm that the binarized templates preserve information about the speaker identity.

At the same time, some inter-speaker pairs demonstrate lower values, for example D2-D6:

$D_H = 23.1\%$. This may indicate partial similarity of binary templates for individual controllers or insufficient stability of some bits after projection-based binarization. This result points to the need for further improvement in stable bit selection and binarization parameter optimization.

6.6. Cryptographic key recovery success

To evaluate the effectiveness of the crypto-biometric module, the successful key recovery rate was determined for each controller. The results are shown in Fig. 7.

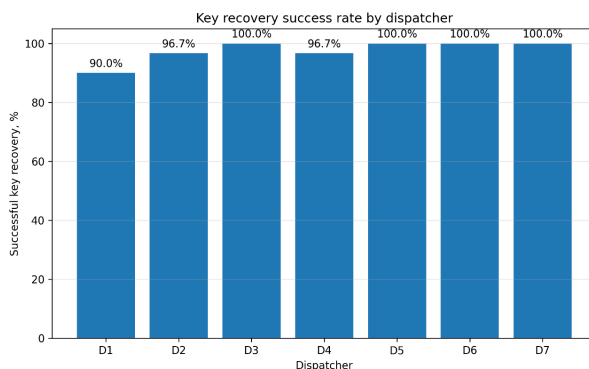


Figure 7: Key recovery success rate by controller

The obtained values show that for most controllers the system provided a very high level of successful key recovery:

$$96.7\% \leq P_{\text{rec}} \leq 100\%.$$

For controllers D3, D5, D6, and D7, a 100% key recovery success rate was achieved. For D2 and D4, the value was 96.7%, and for D1 it was 90.0%.

The lower success rate for D1 may be associated with higher intra-speaker variability, different recording conditions, or less stable bits in the binarized template. This suggests that for some controllers, more enrollment utterances or adaptive selection of binarization and error-correction parameters may be required. Overall, the results confirm the practical feasibility of recovering a cryptographic key from voice biometric data when stabilized binarization and error-correction mechanisms are used.

7. Conclusions

This work developed and experimentally verified a crypto-biometric authentication method for air traffic controllers in digital aeronautical networks based on ECAPA-TDNN, binarization of voice embedding vectors, the Fuzzy Commitment Scheme, and error-correction mechanisms.

Experimental testing on voice samples of controllers D1-D7 showed that ECAPA-TDNN provides high discriminative ability between legitimate and impostor speakers. For comparisons of the same controller, cosine similarity values were mostly in the range 0.73-0.90, whereas most inter-speaker values remained substantially lower.

The constructed ROC curve confirmed the high quality of biometric class separation. This indicates that the proposed approach has high potential effectiveness for voice authentication of air traffic controllers.

At the same time, for aeronautical systems where minimizing the probability of false acceptance of an impostor is critical, it is appropriate to use a stricter operational threshold.

The analysis of binarized templates showed that Hamming distance for samples of the same controller remains low, whereas for different controllers it increases significantly. This confirms that the binarized representation of embedding vectors preserves individual biometric properties and can be used in the Fuzzy Commitment Scheme.

The crypto-biometric module demonstrated a high successful key recovery rate. For most controllers, the recovery rate ranged from 96.7% to 100%, while the minimum value in the experiment was 90.0%. This confirms the principle feasibility of forming a stable cryptographic key based on voice biometric features.

Thus, the experimental results confirm the prospects of using crypto-biometric voice authentication as an additional security layer in digital aeronautical telecommunication networks. The proposed system does not replace PKI, IPsec, or other standard security mechanisms, but it can effectively supplement them by verifying the physical identity of the controller during the establishment or confirmation of a service communication session.

Declaration on Generative AI

The author have not employed any Generative AI tools.

References

- [1] N. Mäurer, T. Gräupl, C. Schmitt, L-Band Digital Aeronautical Communications System (LDACS), RFC 9372, RFC Editor, 2023. URL: <https://doi.org/10.17487/RFC9372>. doi:10.17487/RFC9372.
- [2] S. Brandes, U. Epple, S. Gligorevic, M. Schnell, B. Haindl, M. Sajatovic, Physical layer specification of the L-band Digital Aeronautical Communications System (L-DACS1), in: Integrated Communications, Navigation and Surveillance Conference (ICNS), 2009. URL: <https://doi.org/10.1109/ICNSURV.2009.5172853>. doi:10.1109/ICNSURV.2009.5172853.
- [3] I. C. A. Organization, LDACS SARPs: WP06.AppA-DCIWG-6-LDACS_SARPs, Technical Report, International Civil Aviation Organization, 2023.
- [4] S. J. Undertaking, PJ.14-W2-60 TRL6 final LDACS A/G specification, Technical Report, SESAR Joint Undertaking, 2023.
- [5] Frequentis, The LDACS air-ground datalink, White paper, Frequentis, 2024.
- [6] I. C. A. Organization, LDACS white paper: A roll-out scenario, Information paper, ICAO/ACP DCIWG, 2019.
- [7] O. Holubnychy, et al., Self-organization technique with a norm transformation based filtering for sustainable infocommunications within CNS/ATM systems, in: I. Ostroumov, M. Zaliskyi (Eds.), Proceedings of the 2nd International Workshop on Advances in Civil Aviation Systems Development, volume 992 of *Lecture Notes in Networks and Systems*, Springer Nature Switzerland, Cham, 2024, pp. 262–278. doi:10.1007/978-3-031-60196-5_20.
- [8] EUROCAE, ED-228A: Safety and performance requirements standard for baseline 2 ATS data communications, Technical Report, EUROCAE, 2016.
- [9] I. C. A. Organization, Manual on the Aeronautical Telecommunication Network (ATN) using Internet Protocol Suite (IPS) standards and protocols, Doc 9896, International Civil Aviation Organization, n.d.
- [10] N. Mäurer, A. Bilzhause, A cybersecurity architecture for the L-band Digital Aeronautical Communications System (LDACS), in: 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), 2018. URL: <https://doi.org/10.1109/DASC.2018.8569878>. doi:10.1109/DASC.2018.8569878.
- [11] N. Mäurer, T. Gräupl, C. Schmitt, G. D. Rodosek, H. P. Reiser, Advancing the security of LDACS, IEEE Transactions on Network and Service Management 19 (2022) 5237–5251. URL: <https://doi.org/10.1109/TNSM.2022.3189736>. doi:10.1109/TNSM.2022.3189736.
- [12] M. Vihuro, A. Malyar, G. Litawa, K. Kluczevska-Chmielarz, T. Konrad, P. Migo, Lightweight semantic segmentation for fermentation foam monitoring: A comparative study of u-net, deeplabv3+, fast-scnn, and segnet, Applied Sciences 16 (2026) 1487.
- [13] S. Ivanov, O. Korchenko, G. Litawa, P. Oliinyk, O. Oliinyk, Prototype of hydrochemical regime monitoring system for fish farms, Sensors 26 (2026) 497.
- [14] S. Semenov, V. Tsukur, V. Molokanova, M. Muchacki, G. Litawa, M. Mozhaiev, I. Petrovska, Mathematical model of the software development process with hybrid management elements, Applied Sciences 15 (2025) 11667.
- [15] H. F. Konakhovych, V. P. Babak, V. P. Kharchenko, Y. V. Zaitsev, I. K. Zakharchenska, Method of aircraft navigation with prevention of unauthorized access to flight control in emergency

- situations, Ukraine Patent No. 62506, 2003. URL: <https://iprop-ua.com/inv/zaroccpt/>, national Aviation University. Application No. 2003042938; filed April 3, 2003; published December 15, 2003.
- [16] A. Juels, M. Wattenberg, A fuzzy commitment scheme, in: Proceedings of the 6th ACM Conference on Computer and Communications Security, 1999, pp. 28–36. URL: <https://doi.org/10.1145/319709.319714>. doi:10.1145/319709.319714.
 - [17] Y. Dodis, L. Reyzin, A. Smith, Fuzzy extractors: How to generate strong keys from biometrics and other noisy data, in: C. Cachin, J. L. Camenisch (Eds.), Advances in Cryptology – EUROCRYPT 2004, volume 3027 of *Lecture Notes in Computer Science*, Springer, 2004, pp. 523–540. URL: https://doi.org/10.1007/978-3-540-24676-3_31. doi:10.1007/978-3-540-24676-3_31.
 - [18] Y. Dodis, R. Ostrovsky, L. Reyzin, A. Smith, Fuzzy extractors: How to generate strong keys from biometrics and other noisy data, *SIAM Journal on Computing* 38 (2008) 97–139. URL: <https://doi.org/10.1137/060651380>. doi:10.1137/060651380.
 - [19] U. Uludag, S. Pankanti, S. Prabhakar, A. K. Jain, Biometric cryptosystems: Issues and challenges, Proceedings of the IEEE 92 (2004) 948–960. URL: <https://doi.org/10.1109/JPROC.2004.827372>. doi:10.1109/JPROC.2004.827372.
 - [20] O. Solomentsev, M. Zaliskyi, O. Zuiev, Estimation of quality parameters in the radio flight support operational system, *Aviation* 20(3) (2016) 123–128. doi:10.3846/16487788.2016.1227541.
 - [21] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, S. Khudanpur, X-vectors: Robust DNN embeddings for speaker recognition, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 5329–5333. URL: <https://doi.org/10.1109/ICASSP.2018.8461375>. doi:10.1109/ICASSP.2018.8461375.
 - [22] B. Desplanques, J. Thienpondt, K. Demuynck, ECAPA-TDNN: Emphasized channel attention, propagation and aggregation in TDNN based speaker verification, in: Proceedings of Interspeech 2020, 2020, pp. 3830–3834. URL: <https://doi.org/10.21437/Interspeech.2020-2650>. doi:10.21437/Interspeech.2020-2650.
 - [23] N. Dawalatabad, M. Ravanelli, F. Grondin, J. Thienpondt, B. Desplanques, H. Na, ECAPA-TDNN embeddings for speaker diarization, in: Proceedings of Interspeech 2021, 2021, pp. 3560–3564. URL: <https://doi.org/10.21437/Interspeech.2021-941>. doi:10.21437/Interspeech.2021-941.
 - [24] A. K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, *IEEE Transactions on Circuits and Systems for Video Technology* 14 (2004) 4–20. URL: <https://doi.org/10.1109/TCSVT.2003.818349>. doi:10.1109/TCSVT.2003.818349.
 - [25] I. O. for Standardization, ISO/IEC 24745:2022: Information security, cybersecurity and privacy protection – Biometric information protection, Technical Report, International Organization for Standardization, 2022.
 - [26] I. O. for Standardization, ISO/IEC 30136:2018: Information technology – Performance testing of biometric template protection schemes, Technical Report, International Organization for Standardization, 2018.
 - [27] I. S. Reed, G. Solomon, Polynomial codes over certain finite fields, *Journal of the Society for Industrial and Applied Mathematics* 8 (1960) 300–304. URL: <https://doi.org/10.1137/0108018>. doi:10.1137/0108018.
 - [28] S. B. Wicker, V. K. Bhargava (Eds.), Reed-Solomon codes and their applications, IEEE Press, 1994.
 - [29] K. Hofbauer, S. Petrik, ATCOSIM Air Traffic Control Simulation Speech Corpus, Technical report, 2007.
 - [30] K. Hofbauer, S. Petrik, H. Hering, The ATCOSIM corpus of non-prompted clean air traffic control speech, in: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008), 2008.
 - [31] M. Ravanelli, T. Parcollet, P. Plantinga, A. Rouhe, S. Cornell, L. Lugosch, C. Subakan, N. Dawalatabad, A. Heba, J. Zhong, J.-C. Chou, S.-L. Yeh, S.-W. Fu, C.-F. Liao, E. Rastorgueva, F. Grondin, W. Aris, H. Na, Y. Gao, et al., SpeechBrain: A general-purpose speech toolkit, arXiv, 2021. URL: <https://doi.org/10.48550/arXiv.2106.04624>. doi:10.48550/arXiv.2106.04624.
 - [32] SpeechBrain, Speaker verification with ECAPA-TDNN embeddings on VoxCeleb, Hugging Face model card, 2021. URL: <https://huggingface.co/speechbrain/spkrec-ecapa-voxceleb>.