Cognitive decoding of redundant block codes in the system of processing and protection of biometric data of air passengers

A A Gladkikh¹³, Al K Volkov², An K Volkov², V M Il'in² and D A Kozlov²
¹ Ministry of Education and Science of the Russian Federation, Ulyanovsk State Technical University, Severnyj venec, 32, 432027, Ulyanovsk, Russia
² The Federal Air Transport Agency, Ulyanovsk Civil Aviation Institute, Mozhaisky Street, 8/8, 432071, Ulyanovsk, Russia
³ E-mail: a_gladkikh@mail.ru

Abstract. The article proposes a unified principle of permutation decoding (PD), which applies to redundant systematic block codes. This method allows using correction capabilities of such codes, but in the classical interpretation, the method requires cumbersome matrix calculations, which does not allow using the positive properties of the error correction method. The computing process is excessively complex. Therefore, to reduce the negative effect in the PD system, it is proposed to use the cognitive principle of data processing at the channel level, which significantly reduces the decoder complexity and ensures PD application in the air passenger biometric data systems.

1. Introduction
A key element of the development process of the digital economy of the Russian Federation is the information volume, which basically relies on progressive, cognitive, radiophotonic technological solutions, "big" data, artificial intelligence, virtual and augmented reality methods with widespread use of network technologies. The intensive development of these technologies directly affects the digital transformation of airlines, for which aviation security issues are an urgent task that needs to be addressed through the synthesis of new technological solutions from various subject areas. The highly reliable biometric identification of passenger traffic, the introduction of network technologies based on multimode optical lines and the use of noise-resistant coding for the efficient use of data and their reliable storage have particular importance in this series. This is because the biometric identification procedure is objectively associated with the processing of a large amount of data within a limited time interval. This explains the use of optical lines, which, for economic reasons, cannot be single-mode, and since multimode optical lines have relatively low transfer characteristics of optical signals, thus the problem of protecting data from the influence of destructive factors arises. In addition, the issue of protecting confidential data and system parameters in a data storage system is of fundamental importance.

A key element of the digital economy development in the Russian Federation is the information space, which basically relies on progressive, cognitive, radiophotonic technological solutions, "big" data, artificial intelligence, virtual and augmented reality methods with widespread use of network technologies. The intensive development of these technologies directly affects the digital transformation of airlines, for which aviation security issues are an urgent task that needs to be
addressed through the synthesis of new technological solutions from various subject areas. The highly reliable biometric identification of passenger traffic, the introduction of network technologies based on multimode optical lines and the use of noise-resistant coding for the efficient use of data and their reliable storage have particular importance in this series. This is because the biometric identification procedure is objectively associated with the processing of a large amount of data within a limited time interval. This explains the use of optical lines, which cannot be single-mode for economic reasons, and since multimode optical lines have relatively low optical signals transfer characteristics, thus the problem of data protection from destructive factors arises. In addition, the issue of protecting confidential data and system parameters in a data storage system is of fundamental importance.

The process of air passenger identification consists of two stages. The registration stage, at which user data is recorded, compressed and stored, for example, fingerprint parameters or capture some important features of the subject's face. At the stage of identification and observation, a fingerprint or several fingerprints, the face parameters of the subject are compared with the stored information in the database. After which an affirmative answer is given, and in case of doubt, the subject is monitored for additional parameters. At the same time, control of face parameters is carried out by the non-contact method, which helps to increase the throughput of the control system. This is especially important for large transport hubs, including airports. Since the system includes many users, both tasks associated with the storage and retrieval of the required information are difficult, of course. For this reason, the specified reliability and high speed of data processing in such systems are of fundamental importance.

At the same time, digital passenger identification is limited by possible errors in identification caused by both unintended actions of passengers and deliberate actions by a potential intruder, as well as during biometric data storage and protection. There is also a possibility of network attacks on distributed databases that store air passenger biometric containers. Inadvertent actions are associated with the stochastic nature of human biometric data and reflect the fact that these characteristics change over time. Deliberate actions reflect simulated attacks when an unauthorized person changes his biometric profile to look like an authorized passenger. The effective method to eliminate the influence of the considered negative factors is to apply error-correcting coding algorithms. Some results of using such algorithms in biometric identification systems are presented in [1-3]. However, the existing algorithms do not fully possess the required performance that meets the conditions of large passenger traffic and adaptability to the problems of storing and protecting biometric characteristics in data warehouses.

Thus, the choice of an effective error-correcting coding algorithm is one of the most important elements in prospective systems of air passenger biometric identification.

2. Materials and methods
The basis of the PD method is the property of linear block systematic redundant codes to maintain their weight structure for various sets of information symbols with length of \( k \). Fundamentally, permutations of the elements of the code combination of length \( n > k \) are carried out on the basis of a particular concept. For example, priority is given to symbols that have the highest gradations of reliability ratings. The closer the accepted value of the soft decision to the mathematical expectation of such estimates, the higher the index of the soft decision of the symbol., It is possible to form a combination of an equivalent code (EC) by rearranging the symbols with the most reliable estimates on the positions of information bits on the receiving side and coding them. The weight of such a combination will be adequate to the weight of the source code combination, and if reverse permutations are performed in the EC combination, the combination of the source code and the combination of the rearranged code will exactly match. Bitwise addition of such combinations in the absence of interference will provide a zero vector.

The relevance of such a transition is emphasized by the possibilities of using permutation decoding in modern coherent networks, which are characterized by contradictions with respect to more and more increasing data exchange rates in optical systems, with some technology lagging in the data processing speed in the codec processors of such systems. At the same time, it should be pointed out
that in promising coherent networks there is a tendency to use non-binary corrective codes, which necessitates the study of such codes to solve problems of increasing the spectral and energy efficiency of these systems. In this regard, we consider the technology of permutation decoding of linear systematic codes using the example of the non-binary Reed-Solomon (RS) code in more detail.

In the most general case, the permutation decoding process for block codes can be represented by the expression

\[ P_d(t) = \begin{cases} Rn(t) = I(t) + Er(t); \\ D(t). \end{cases} \]  

(1)

The following notation is introduced in expressions (1): \( Rn(t) \) – a random component of the decoding procedure, which includes the random process \( I(t) \) of selecting a code word semantically corresponding to the transmitted message as the source of information and the random process \( Er(t) \) of the influence of interference in the communication channel used by the data transmission system at the physical level. The parameter \( D(t) \) is the deterministic component of the decoding process, which is expressed in connection with a specific permutation of characters from the corresponding EC. Suppose that during the implementation of procedure (1), during the operation of the decoder, at the moment of time \( t_1 \), a rearrangement of column numerators of the form is formed \( perm(t_1) \), and for such a rearrangement, the decoder performed all the actions associated with the search for the inverse matrix \( Q^T(t_1) = M_Q^T/\Delta \) where \( Q \) – a matrix of columns at positions \( k \) of the most reliable elements, \( M_Q^T \) – transposed matrix of matrix minors \( Q \) and \( \Delta \neq 0 \) is determinant of this matrix. The indicated procedure is key for the search for generators of matrices of EC \( G_{sys}(t_1) \), at the same time, time was spent on the production of calculations \( t_1 \). Subsequently, when transmitting data, a situation may arise when the rearrangement of column numerators can be repeated at some arbitrary and discrete time \( t_i \).

In accordance with the classical concept of PD systematic block codes, the decoder re-performs the search \( G_{sys}(t_1) \) at the first stages of data processing operates in accordance with the classical concept, but remembers the results of processing the current permutations of column numerators and flexibly uses these results as they accumulate in RAM, forming a region \( D(t) \).

Several scientific problems arise against this background:

- whether it is possible to remember the result of calculating the matrix \( G_{sys}(t_1) \) in the RAM of the receiver processor and use it in case of repeated permutation of the column numerators at the i-th time moment due to the formation of the component \( D(t) \);
- whether it is possible to use cognitive data processing methods in the receiver decoder when the permutation decoder at the first stages of data processing operates in accordance with the classical concept, but remembers the results of processing the current permutations of column numerators and flexibly uses these results as they accumulate in RAM, forming a region \( D(t) \);
- whether it is possible to obtain the results of permutations of column numerators due to external computing resources and their introduction into the decoder memory at the stage of its preliminary preparation for work;
- whether it is permissible to store in the RAM of the receiver processor the entire volume of possible results of evaluating permutations of column numerators;
- how the memory of storing the results of permutations should be organized in order to quickly and efficiently search for the permutation required at the current moment.
It was found that all of the above problems have a positive solution, which provides a significant gain in data processing speed in the decoder by replacing complex matrix calculations with a certain amount of memory required for the cognitive decoder card (CCD). We show this using code as an example RS (7, 3, 5).

Let the selected numbers of reliable characters in some accepted vector of the RS code be positions 2 5 4 and let the numerators of less reliable characters make up the sequence 3 6 7 1. Then the generating matrix of the code takes the form, while the lower indices denote the column numerators, and the upper indices denote the degrees of the primitive element of the field $GF(2^3)$.

$$G = \begin{pmatrix}
\alpha_1 & 0 & 2 & \alpha_4^2 & \alpha_5^0 & \alpha_6^4 & \alpha_7^5 \\
0 & \alpha & 0 & \alpha^2 & \alpha_0 & \alpha^6 & \alpha^6 \\
0 & 0 & \alpha & \alpha^3 & \alpha^0 & \alpha^1 & \alpha^3
\end{pmatrix} \Rightarrow G_{\text{perm}} = \begin{pmatrix}
0 & \alpha_5^0 & \alpha_4^2 & 0 & \alpha_6^4 & \alpha_7^5 & \alpha_1^0 \\
\alpha^0 & \alpha^0 & \alpha^2 & 0 & \alpha^6 & \alpha^6 & 0 \\
0 & \alpha^0 & \alpha^3 & \alpha^0 & \alpha^1 & \alpha^3 & 0
\end{pmatrix} \quad (2)$$

The rearranged matrix from expression (2) can be easily translated into a systematic form.

$$G_{\text{sys}}^{\text{perm}} = \begin{pmatrix}
\alpha^0 & 0 & 0 & \alpha_5^2 & \alpha_6^6 & \alpha_7^2 & \alpha_1^6 \\
0 & \alpha^0 & 0 & \alpha^2 & \alpha^6 & \alpha^4 & \alpha^4 \\
0 & 0 & \alpha^0 & \alpha^1 & \alpha^3 & \alpha^3 & \alpha^1
\end{pmatrix}. \quad (3)$$

It was established by checking that the fourth column from the composition of numbering $r$ of expression (3) with an unchanged sequence and any sequence of numbering from always corresponds to the third numbering. The fifth column will correspond to numbering 6, and the remaining columns will correspond to numbers 7 and 1. From this it follows that changing the order of the numbering in this part of the matrix during linear transformations does not change the structure of the columns, and this factor is used to save computational resource.

The sequence of numerators from the set of reliable characters corresponds to permutations of the rows of the check part of the matrix in expression (3). The idea arises of introducing a special operator, which we call the operator of fast matrix transformations (FMT).

Therefore, for many permutations $k!r!$, there is only one sample of the reference matrix, which should be stored in the CCD. During the operational work of the decoder, to quickly find the required standard and quickly convert it, the numbering of the rows and columns of the standard matrix should be stored in the memory of the CCD in lexicographic format.

Let the reference sample with numerators of the form 2 4 5 and 1 3 6 7 be stored in the CCD. It is noticeable that the lexicographical numerators are ordered. The search for a reference matrix in a systematic form after exclusion of a single matrix from its composition has the form:

$$\begin{pmatrix}
\alpha^6 & \alpha^2 & \alpha^6 & \alpha^2 & 2 \\
\alpha^1 & \alpha^1 & \alpha^3 & \alpha^3 & 4 \\
\alpha^4 & \alpha^5 & \alpha^5 & \alpha^4 & 5^* \\
1 & 3 & 6 & 7
\end{pmatrix} \quad (4)$$

In the expression (4) the numbering of lines and columns are listed separately from matrix elements for extra convenience. The reference matrix in the memory of the CCD allows you to get the matrix of the EC in just three steps. Actually: calling the matrix from the memory of the CCD according to the generated set of numbering, then permutation of the rows and permutation of the columns. From the expression (4) and the numbering sets in the format 2 4 5 and 1 3 6 7 we obtain the sequence of actions...
\[ \begin{align*}
\alpha^6 & \quad \alpha^2 \\
\alpha^1 & \quad \alpha^1 \\
\alpha^4 & \quad \alpha^5 \\
\alpha^3 & \quad \alpha^3 \\
\alpha^4 & \quad \alpha^5 \\
\alpha^3 & \quad \alpha^3 \\
\alpha^2 & \quad \alpha^2 \\
\alpha^6 & \quad \alpha^6 \\
\alpha^2 & \quad \alpha^2 \\
\alpha^6 & \quad \alpha^6 \\
\alpha^2 & \quad \alpha^2 \\
\alpha^6 & \quad \alpha^6 \\
\alpha^2 & \quad \alpha^2 \\
\alpha^6 & \quad \alpha^6 \\
\alpha^2 & \quad \alpha^2 \\
\alpha^6 & \quad \alpha^6 \\
\alpha^2 & \quad \alpha^2
\end{align*} \tag{5} \]

The result obtained corresponds to the result obtained above using the classical procedure for generating the generating matrix of the equivalent code. A comparative gain in the number of operations in the implementation of matrix calculations and the proposed method is shown in Table 1.

| Cod RS (7, 3, 5) | Cod RS (15, 3, 11) | Cod RS (15, 9, 7) | Cod RS (15, 13, 3) |
|-----------------|-------------------|-----------------|-------------------|
| The number of arithmetic operations in the implementation of matrix calculations | 336 | 2410 | 2912994 | 6858434026 |
| 6720·10^{-9} sec. | 48200·10^{-9} sec. | 58259880·10^{-9} sec. | 1371686680520·10^{-9} sec. |
| The number of row and column permutations in the implementation of FMT | 7 | 15 | 15 | 15 |
| 140·10^{-9} sec. | 300·10^{-9} sec. | 300·10^{-9} sec. | 300·10^{-9} sec. |
| The gain in the number of operations | 48 times | 1.6·10^2 times | 9.7·10^3 times | 2.3·10^8 times |

It is noticeable that in order to obtain the desired matrix from the reference form, only \( n \) steps were taken that were not related to performing arithmetic operations in the Galois field \( GF(2^3) \). The method of FMT significantly improves the situation of reducing the computational complexity of the receiver processor. The revealed regularities make it possible to reduce the complexity of the decoder implementation from a value of \( O(n^3) \), typical of the vast majority of matrix calculations, to a value of \( O(n) \) complexity, corresponding to linear data transformations. Assessment of the required amount of memory card was carried out relative to the capabilities of the programmable logic device (PLD) "Altera", the amount of which was taken as 100%. A significant gain in the amount of required memory of the CCD is given by taking into account the cyclic properties of codes RS. The results of the analysis are shown in Table 2.

| Cod | Complete set of matrices | FMT with a system of reference matrices | Consideration of cyclic properties of reference matrices |
|-----|--------------------------|---------------------------------------|------------------------------------------------------|
| RS (15,11,5) | 5715950% | 1.5% | 0.1% |
| RS (15,7,9) | 43300% | 8.6% | 0.5% |
| RS (255,251,5) | Very large volume | 625% | 15% |
| RS (40,36,5) | | | |

The obtained results allow us to state that the use of PD based on the principles of FNT uniquely provides protection and data processing at the required time intervals. At the same time, it should be noted that permutation decoding can significantly increase the reliability of data storage and the efficiency of their recovery in case of loss, as described in [5].

3. Conclusion

This paper describes a new concept for organizing passenger flows at airports based on digital identity and presents its main components. The advantages and disadvantages of this concept are revealed. A
new approach is proposed to increase the reliability of the operation of digital passenger identification systems in terms of storing biometric data while speeding up the processing of digital data.

Reference
[1] Hao F, Anderson R, Daugman J 2006 Combining Crypto with biometrics effectively IEEE Transactions on Computers 55(9) 1081-8
[2] Alawi A Al-Saggaf 2018 Secure Method for Combining Cryptography with Iris Biometrics Journal of Universal Computer Science 24(4) 341-56
[3] Peng L, Xin Y, Hua Q, Kai C, Eryun L and Jie T 2012 An effective biometric cryptosystem combining fingerprints with error correction codes Expert Systems with Applications 39 6562-74
[4] Gladkih A A, Al-Tamimi T F H 2018 The concept of cognitive data processing in a permutation decoding system of a non-binary redundant code Telecommunication 9 69-74
[5] Ganin D V, Gladkikh A A and Klimov R V 2019 Method of organizing storage of data based on codes of products with simple parity check with offset 2699678 vol 25 (Moscow: Rospatent)