The concept for biometric system development based on modern error correcting coding

A A Gladkikh¹,³, Al K Volkov², An K Volkov², B A Saeed¹ and V V Yudaev²

¹Ulyanovsk State Technical University, Severnyj venec, 32, 432027, Ulyanovsk, Russia
²Ulyanovsk Civil Aviation Institute, Mozhaisky Street, 8/8, 432071, Ulyanovsk, Russia
³E-mail: a_gladkikh@mail.ru

Abstract. The objective of the work is to develop new algorithms for error-correcting coding in biometric systems that best meet modern challenges for such systems. The article presents directions of biometric system development that take into account the prospects of their use in combination with modern network technologies, methods of cognitive data processing, cryptographic systems and artificial intelligence. The objective necessity for developing advanced error-correcting coding methods in such systems is proved that meets the increased requirements for the reliability of the processed data and its processing speed in real time. It is proposed to use a cascade design for processing block redundant codes with a system for cluster processing of polar codes (PC) at the internal decoding stage and cognitive processing of Reed-Solomon (RS) code combinations at the external decoding stage.

1. Introduction
The uniqueness of biometric data causes a rapidly growing demand for such information in modern telecommunication systems, cyberphysical systems and personality identification systems. At present individual achievements in this area are increasingly integrated to solve problems in which an increased level of processed data reliability is required, and this necessitates the use of error correcting coding associated with the theory of groups, rings and fields. Among various physiological characteristics which make possible fast and reliable remote biometric identification of people in various real-time processes, the most interesting thing is the facial image. First of all, these tasks include identifying subjects in a dynamic human flow, identifying passengers in transport systems as part of the profiling procedure, solving cryptographic issues in public key systems for correct access to information flow parameters in cyber-physical systems. In the context of globalization, biometric data is increasingly processed based on network process chains which have the stochastic level of interfering and destructive factors. For this reason, the urgent task in processing this information is to protect data from interference on the basis of error-correcting coding, which is clearly long-term taking into account the peculiarities of post-quantum cryptography applied to biometric databases.

2. Analysis and directions of biometric systems development
The traditional access to digital data (DD) is a bunch of "login-password". The combination of letters and numbers is the most common type of password used by users of various systems. However,
modern knowledge and abilities of attackers allow them to decrypt such a password easily or use an attack by direct brute force. The next type of authentication is identification using cryptographic methods that use long passwords to organize strong protection. Often, due to the weak protection of passwords themselves, their effectiveness remains low. These authentication methods have many problems associated with the possibility of theft, loss, copying or transferring keys and tokens to others. Therefore, as technologies develop to solve the problem of user authentication, methods based on measuring human biometric parameters are actively developing. This is due to the fact that copying human biometric data for violators involves more labor than copying user passwords. In addition, biometric data cannot be forgotten or lost, and biometric identification itself assumes the physical presence of the user at the entry point into the system. The presented type of biometric systems is a weak (ordinary) biometry. Besides natural errors, which are associated with the fact that individual biometric samples obtained during registration and authentication are fuzzy, ordinary biometric systems are vulnerable to direct attacks by attackers on various system modules.

To exclude the possibility of unauthorized access to DD, it is proposed to use biometric cryptography, which involves "linking" encryption keys and passwords with the biometric parameters. Unlike traditional biometric identification in these systems, each biometric template is pre-converted into a bit (key) sequence, which is used to authenticate the subject. In this case, the biometric template is stored in the form of auxiliary information that does not allow recovering biometric parameters of the subject. The main component of biometric cryptographic systems is the biometrics-code converter, which serves to convert the vector of fuzzy biometric parameters "one’s" into a clear unambiguous code of the access key (password). The most common biometrics-to-code converter is a fuzzy extractor, which is a method that extracts random, evenly distributed sequences of bits from biometric data in noisy environments [1]. The development of this approach is "fuzzy vault" [2] and "fuzzy commitment" [3]. A key element of these approaches is error correction codes. They apply to biometric data to correct unstable bits of a generated password key.

Based on the experience of using biometric systems based on fuzzy extractors, the following requirements for noise-resistant coding schemes could be distinguished:

1) firstly, the scheme should provide distinguishability of variability within a class and between classes of biometric data by improving the corrective ability. This requirement is directly related to errors of the first and second kind. To minimize such errors, it is necessary to identify distances that distinguish one class of biometric data from others. It is on the basis of this data that it is advisable to choose a class of error-correcting code with one or another correcting ability and complexity of the implementation of the decoding procedure;

2) secondly, the circuit must be high-speed in order to ensure good fuzzy extractor performance. This is especially important when processing a large amount of biometric information in real time;

3) thirdly, the scheme should have a tool for regulating the relative speed of the code to achieve a compromise between the errors of the first and second kind. This implies adaptive algorithms for regulating relative speed of the code, the use of cognitive data processing methods and the use of developed artificial intelligence systems.

The last two positions require their scientific permission, both in theoretical and practical terms.

3. Perspective data protection algorithms from influence of stochastic interference

Among numerous means of error-correcting coding, the sequential (cascade) and parallel turbo coding algorithms are most demanded. The latter require very long code combinations and are not technologically adapted to obtain soft symbol solutions (SSS) in modern optical communication channels. The implementation of the "belief propagation" method, which is the basis for this algorithm, is impossible without SSS. For this reason, the cascade coding method is more suitable for solving data protection problems regarding a limited amount of digitized parameters of human faces. At the same time, it is advisable to use PC at the internal stage of the codec system.

The technology for processing PC relies on the conversion of a continuous communication channel into a system of vector channels with cross-connections and a complete exclusion of the accepted
sequence of channels which transmit bits obviously unreliably. The bandwidth of such channels is considered to be equal to zero (channels are considered "frozen"), and data recovery is carried out due to information obtained from reliable channels. The purpose of using PC is to create such a crucial scheme in which the probability of erroneous decoding of code vector \( Y(P) \rightarrow 0 \), thereby the probability of erroneous reception also tends to zero. The concept of constructing a PC is based on the Arikan kernel, which is represented in the form of a matrix \( F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \), while \( F^{\otimes m} \) denotes its \( m \)-th Kronecker degree, which is based on the direct matrix product [4]. Thus, a suitable code for the inner stage can be obtained. To obtain the required output vector, it is necessary to transform the sequence so that the number of the new position of the \( i \)-th element is obtained as the inverse record of the number \( i \), i.e. polar representation. For example, \( 1 \equiv (0001) \rightarrow (1000) \equiv 8 \). To obtain such transformations, a permutation matrix \( B_n \) is introduced, then the generating matrix of the code takes the form \( G^{\otimes m} = F^{\otimes m} \times B_n \) and the code is considered to be given. SSS for PC are obtained as Bhattacharia coefficients \( Y = \sqrt{P(z|0) \cdot P(z|1)} \), where \( P(z|0) \) and \( P(z|1) \) are the corresponding probability distributions of binary signals, which are stable enough for optical communication channels and can be detected a priori. This provides an indirect SSS implementation in optical communication channels.

At the external stage of the codecs, it is advisable to use the RS code, which leads to reliable identification in conditions of both simple pairwise comparisons and in conditions of more complex biometric data. A similar approach is reflected in [5]. It uses a direct product of the Hadamard code (64,7,32) and the RS code (32,20,13) over GF \( (2^7) \). However, the effectiveness of using this code combination is not completely clear. The proposed cascade design in combination with permutation decoding methods at the external data processing stage, is able to ensure maximum use of the redundancy introduced into the code. To a large extent, the overall error correction result depends on the capabilities of the code in the inner stage. In optical systems (whether it is single-mode or multi-mode), to increase the data transfer rate, complex types of modulation are used, such as QPSK or PAM-4, which, providing a high data exchange rate, objectively reduce their reliability. This leads to the fact that the probability of an error per bit in such systems decreases to level \( 10^{-5} \) and, due to the code of the external stage, increases to the required probability of \( 10^{-12} \). However, external stage codes require large time intervals that can be shortened by cognitive procedures.

Practically no data transmission system using redundant coding allows the decoder training procedure to achieve the effect of accelerating the received data. The receiver, with rare exceptions, is not able to predict the behavior of the information source, which can send over the channel with interference of any of a given set of combinations \( I(t) \). In addition, a priori, an unknown interference vector \( Er(t) \) acts in the communication channel, which together with \( I(t) \) increases the uncertainty for the receiver. The only method of data processing at the reception for which it is possible to specify the deterministic component \( P(t) \), is the method of permutation decoding (PD), described in [6]. Therefore, in the most general case, the PD process for block codes can be represented by the expression:

\[
PD = \begin{cases} 
Rn(t) = I(t) + Er(t), \\
Dn = P(t), \\
S_1, S_2, S_3_{\text{max}}
\end{cases}
\]

where \( S_1 \) – preset level of impulse synchronization, \( S_2 \) – required level cycle synchronization, \( S_3 \) – set level synchronization by key parameters of the cryptographic system.
The presence of the deterministic component \( P(t) \) (1) allows the decoder to "train" to recognize specific permutations and provide ready-made data processing decisions depending on the current situation in selecting reliable characters. Learning outcomes are usually stored in the memory of the decoder, which is implemented in the format of the cognitive card of the decoder (CCD). The CCD is considered perfect provided that all permissible permutations are introduced into it. If the card is not perfect, then the decoder may be in training mode, when for the known permutations, the finished result is extracted from the CCD, and if the permutations are not known, this result is calculated and stored in the corresponding memory cell of the CCD. It becomes clear that after a certain time an imperfect CCD becomes perfect.

The memory of the CCD is organized lexicographically for quickly finding the desired result. The structure of such a map is described in [6]. It is known from combinatorics that permutations are subject to cyclic shifts, the organization of which is carried out using orbits. The study found that each lexicographically ordered generatrix combination of any cycle among all orbits for a given code corresponds to a unique and different verification matrix of the group code, which is part of the reference generating matrix of the equivalent code for the corresponding set of numerators after substituting to the verification matrix corresponding to dimensions of the identity matrix on the left. With a cyclic shift of the numbering of the generatrix in a lexicographically ordered form to the right, only the cyclic shift of the columns of the reference check matrix to the right occurs, provided that the largest of the valid numbering on the right does not turn into the smallest numbering on the left. With a cyclic lexicographical shift of the generatrix numerators to the right, it is occurred a rows cyclic shift of the reference verification matrix from bottom to top, provided that the largest of the valid numbers on the right when moving along the cycle turned into the smallest numerator on the left. In cases different from the lexicographically ordered sequence of permutation numerators, the orbit reference matrix is reduced to the required verification matrix by sorting the columns of the reference matrix in the order determined by the specific permutation of information bits, and the sorting of rows due to the sequence of numerators is determined by the permutation of the checking bits.

Using the results of the above PD procedure we can drastically reduce the data processing time, which is positively reflected in the conditions of high-speed single-mode or multimode optical communication lines when using complex types of modulation.

4. Conclusion

Thus, PC in data exchange systems allows to implement cluster decoding algorithms and providing a flexible system for protecting biometric data in modern DD exchange systems, and also to develop technologies for future promising systems of this kind.

Acknowledgments

The reported study was funded by RFBR and INSF, project number 19-47-730011.

Reference

[1] Dodis Y et al 2008 Fuzzy extractors: How to generate strong keys from biometrics and other noisy data SIAM Journal on Computing 1(38) 97-139
[2] Juels A, Sudan M 2006 A fuzzy vault scheme Designs, Codes and Cryptography 2(38) 237-57
[3] Juels A, Wattenberg M 1999 A fuzzy commitment scheme Proceedings of the 6th ACM Conference on Computer and Communications Security pp 28-36
[4] Gladkikh A A, Mishin D V and Chilikhin N Y 2018 Improving efficiency of fiber optic communication systems with the use of lexicographic decoding of polar codes Proceedings of SPIE. Optical Technologies for Telecommunications 10744 1-15
[5] Hao F, Anderson R, Daugman J 2006 Combining Crypto with biometrics effectively IEEE Transactions on Computers 55(9) 1081-8
[6] Gladkikh A A, Volkov An K, Volkov Al K and Ibragimov R Z 2019 Noiseless coding algorithm based on cognitive processing of biometric data in system of digital identification of
passengers \textit{Journal of Physics: Conference Series (ITBI 2019)} \textbf{1333} 1-5