UDC 003.26
doi:10.31799/1684-8853-2022-3-45-54

The variant of post-quantum cryptosystem based on burst-correcting codes and on the complete decoding problem

A. A. Ovchinnikov[a]***,*** *PhD, Tech., Associate Professor, orcid.org/0000-0002-8523-9429, mldoc@guap.ru*
[a]*Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaia St., 190000, Saint-Petersburg, Russian Federation*

Introduction: *Today the investigations of post-quantum cryptosystems secure against quantum computations is the area of great interest. An important direction here is code-based cryptography utilizing the mathematical problems from error-correcting coding theory. The improvement of existing code-based systems may be achieved both in practical part (reducing the key sizes) and theoretically by using more complicated mathematical code-based tasks.* ***Purpose:*** *The development of public-key code-based cryptosystem using low-density parity-check codes with burst correction; the estimation of the parameters of the obtained system.* ***Results:*** *The variant of code-based cryptosystem using random block permutation low-density parity-check codes is proposed. The cryptocomplexity of the system is supposed to be based on the complete decoding problem, which is believed to be a harder mathematical problem than those used in existing systems. With high probability, the analysis of the system by using decoding methods is not possible at all, which both increases the long-term cryptocomplexity of the system and allows to reduce the key size. The evaluation of the underlying code selection is performed, the approaches to the selection of the parameters of the proposed system on the basis of the required level of cryptocomplexity are considered.* ***Practical relevance:*** *The proposed system allows to reduce the public-key size as compared to the classical McEliece system, cryptocomplexity also comparable, with the underlying mathematical problem to be more stable against perspective attacks.*

Keywords *— post-quantum cryptography, code-based public-key systems, low-density parity-check codes, burst error correction.*

For citation: Ovchinnikov A. A. The variant of post-quantum cryptosystem based on burst-correcting codes and on the complete decoding problem. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2022, no. 3, pp. 45–54. doi:10.31799/1684-8853-2022-3-45-54

Introduction

The concept of public-key cryptography is usually connected with groundbreaking paper by W. Diffie, M. Hellman “New directions in cryptography” published in 1976 [1]. According to this concept each part has the pair of long-term keys: public key and correspondent private (secret) key. In case of secrecy providing the recipient’s public key is used during encryption, while the correspondent private key is used for decryption. Today the most widely spread public-key system is RSA whose strength is based on hardness of integer factorization. However, this problem is not belongs to NP-hard problems [2], besides, the quantum polynomial-time Shor’s algorithm is known for this task, so in middle-term perspective the strength of RSA becomes under question both in terms of classical computation architectures and by using powerful enough quantum computers. It worth to mention that there are intensive arithmetic with big integers (order of thousands of bits) being used in RSA system, so practical implementations of this system are rather slow.

As the result in 2016 NIST initiated the competition on adoption the new post-quantum cryptography standard [3]. One of the main directions within post-quantum cryptography is code-based cryptography, utilizing the problems from error-correcting codes theory.

The first code-based cryptosystem was proposed by R. McEliece in 1978. Being extremely computationally efficient, McEliece system, nevertheless, did not found wide practical usage, which is traditionally explained by relatively large key sizes, primarily for public key. Possible directions of McEliece system improvement are usage of error-correcting codes classes allowing decreasing the public key size, as well as selection of more complicated mathematical problems for system’s strengthening.

In this paper, the public-key system based on specific class of error-correcting low-density parity-check codes for bursts error correction is considered. The system uses the hard problem of complete decoding, which is NP-hard and potentially harder than the bounded-distance decoding problem used in McEliece cryptosystem.

Code-based hard problems

For investigating and understanding the details of different code-based cryptosystems the basics of underlying hard problems should be considered.

Public-key cryptography is based on the concept of one-way trap-door functions. Briefly the construction of such functions may be described as follows:

— (P, S) — key pair, where P — public key, S — private (secret) key;

— $E_P(m) = c$ — polynomial-time function, mapping the message m into ciphertext c using P;

— $D_S(c) = m$ — polynomial-time function, which is inverse to E, and uses S.

From the point of view of cryptographic strength the following should be provided:

— with knowledge of P, calculation of S should be computationally hard;

— with knowledge of c, and without knowledge of S, calculation of $E^{-1}(m) = c$ should be computationally hard.

By computational hardness it is supposed the exponential-time complexity of the correspondent problem, however, the specificity of one-way trap-door functions is that their inversion should be hard in general, but feasible (polynomial-time) with knowledge of secret S.

In the complexity theory there are approaches for problems classification by so-called “feasible” and “hard”, one of the most widely used approach considers the following classes:

— class P (polynomial) — problems which may be solved by polynomial time on deterministic Turing machine;

— class NP (non-deterministic polynomial) — problems which may be solved by polynomial time on non-deterministic Turing machine (note $\mathrm{P} \subseteq \mathrm{NP}$);

— class NPC (NP-complete) — problems which are in NP, and any other problem from NP can be reduced to them by polynomial time;

— class NP-hard — problems which may be not from NP, but any other NP-complete problem can be reduced to them by polynomial time.

More formal and accurate mathematical definitions are out of the scope of this paper, in cryptography the NP-hard problems are usually considered, but we will not make distinction between NP-complete and NP-hard problems.

Within this classification the problems from P are considered as feasible, while NPC or NP-hard contain hard problems (for which only exponential-time solutions are known in general case), however, polynomial-time specific cases are possible. We should also mention the existence of problems (denote them as “< NPC”, which means “hard problems but simpler than NPC”), for which the polynomial-time solution in general case is unknown, but these problems are simpler than NPC in the sense that if their polynomial solution would be found this will not help to solve the problems from NPC. For example, such problems are integer factorization or discrete logarithm problem that are used in most practically spread number-theoretic cryptosystems.

Consequently, the following classes may be used to construct the one-way trap-door functions:

— < NPC — widely used in cryptography for today, but it is believed that there are the possibility of finding the polynomial-time solutions for these problems, besides, number-theoretic problems from this class have quantum polynomial complexity (may be solved by polynomial time using quantum computer);

— NPC — it is believed that polynomial solution for this class do not exist at all (though this is not proved mathematically), there are no polynomial time quantum algorithms known for this class.

From the classification given above it follows that NP-complete (or NP-hard) problems are preferable for usage in cryptography, but the distinction should be made between the cryptosystem (i.e. trap-door function) and underlying hard mathematical problem — it may be turned out that trap-door function does not belong to the same class as correspondent hard problem, for example the Merkle — Hellman system was broken by A. Shamir by polynomial time [1], though correspondent subset-sum problem is NP-hard.

Next we describe several hard problems from the coding theory, for this goal some definitions and terms should be given.

Linear (n, k)-code is k-dimension subspace of n-dimension linear vector space over the field F (in this paper only binary codes over $GF(2)$ are considered) [4, 5]. We assume $k < n$, then k is the number of information symbols, n is codelength, the value $r = n - k$ defines the number of redundant symbols, $R = k/n$ is code rate. Since linear code is linear vector space, it may be defined by its basis $\mathbf{G}$, which is $(k \times n)$-matrix called the generator matrix of the code. Basis of the orthogonal space is $(r \times n)$-matrix $\mathbf{H}$, which is called the parity-check matrix of the code, and $\mathbf{GH}^{\mathrm{T}} = \mathbf{0}$. If $\mathbf{m}$ is k-bit information vector, then $\mathbf{a} = \mathbf{mG}$ is codeword of length n, the vector $\mathbf{S} = \mathbf{bH}^{\mathrm{T}}$ is called the syndrome for arbitrary vector $\mathbf{b}$ of length n, and $\mathbf{S} = \mathbf{0}$ iff $\mathbf{b}$ is codeword.

Let C is the set of codewords, $\mathbf{a} \in C$ — any codeword of length n, $\mathbf{b}$ is arbitrary vector of length n. The difference between $\mathbf{b}$ and $\mathbf{a}$ may be described by the so-called error vector $\mathbf{e} = \mathbf{b} - \mathbf{a}$ (we assume binary arithmetic which uses XOR), or $\mathbf{b} = \mathbf{a} + \mathbf{e}$.

The problem of minimal distance decoding is an optimization problem

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}\in C} d(\mathbf{a},\mathbf{b}), \tag{1}$$

where $d(\mathbf{a}, \mathbf{b})$ is Hamming distance between $\mathbf{a}$ and $\mathbf{b}$.

The problem of bounded-distance decoding, or decoding in sphere with radius t is an optimization problem (1) with additional constraints:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}\in C, d(\mathbf{a},\mathbf{b})\le t} d(\mathbf{a},\mathbf{b}). \tag{2}$$

Note that the solution of (2) is not always exists, and $d(\mathbf{a},\mathbf{b}) = W(\mathbf{b} - \mathbf{a}) = W(\mathbf{e})$, where $W(\mathbf{e})$ is Hamming weight of $\mathbf{e}$.

The minimal distance d_0 of the code is the minimal pairwise Hamming distance between codewords. Then the code can correct any combination of t errors or less, where $d_0 = 2t + 1$, this means that if no more than t symbols are incorrect in codeword $\mathbf{a}'$, i.e. $\mathbf{b} = \mathbf{a}' + \mathbf{e}$, $W(\mathbf{e}) \le t$, then the problem (2) of bounded-distance decoding in sphere with radius t always has exactly one solution, which is $\hat{\mathbf{a}} = \mathbf{a}'$.

Linear (n, k)-code split the overall n-dimensional vector space into 2^r disjoint sets, one of them is the set of codewords and others are cosets. All vectors from the coset has the same syndrome (which is zero vector for the set of codewords). From any coset one representative may be chosen which is called the coset leader (zero codeword for the set of codewords). Since there are 2^r leaders and also 2^r different syndromes, the one-to-one mapping may be set between them, allowing to define the syndrome decoding procedure as calculation the syndrome $\mathbf{S} = \mathbf{bH}^{\mathrm{T}}$ for the vector $\mathbf{b}$, then the leader of the coset with correspondent $\mathbf{S}$ is considered as error vector $\mathbf{e}$, and the decoded codeword is $\hat{\mathbf{a}} = \mathbf{b} - \mathbf{e}$. If the coset leader is chosen as the vector with minimal weight from the coset, then syndrome decoding coincides with minimal distance decoding [4–6].

Note that in fact the list of coset leaders coincides with the set of errors correctable by the code. We will call the decoding, allowing correction of any coset leader, as complete decoding. Clearly, bounded-distance decoding is incomplete: only the subset of leaders with weight of no more than t may be corrected.

For the random linear code it is proved that the following problems are NP-hard:

— minimal distance decoding;

— complete decoding;

— calculation of the code's minimal distance;

— calculation of the non-zero codeword of minimal weight.

Note that the bounded-distance decoding problem is not in the list, though there are different points of view concerning NP-hardness of this problem, however, to the author's knowledge, formal proof of any correspondent hypothesis is unknown. It should be mentioned that the listed problems are hard for random linear codes, while for some specific code constructions simple solutions are known, this allows usage of coding problems in construction of public-key cryptosystems.

Classical code-based public-key cryptosystems

The idea of the McEliece system [1, 7, 8] is to select the error-correcting code, for which effective (polynomial-time) decoding algorithm is known, and to hide the structure of this code in linear code of random structure. This idea is realized as follows.

1. Key generation.

Each entity U performs the following.

— Select generator $(k \times n)$-matrix $\mathbf{G}$ of linear code, which can correct t errors (has minimal distance $d_0 \ge 2t + 1$), and for which the polynomial-time bounded-distance decoding procedure ψ is known (in the sphere of radius t).

— Compute $\mathbf{G}' = \mathbf{MGP}$, where $\mathbf{M}$ — non-singular $(k \times k)$-matrix, $\mathbf{P}$ — $(n \times n)$-permutation matrix.

— Public key is $P_{\mathrm{U}} = (\mathbf{G}', t)$, private key is $S_{\mathrm{U}} = (\mathbf{M}, \mathbf{G}, \mathbf{P})$.

2. Encryption.

Entity A encrypts k-bit message $\mathbf{m}$, using authentic public key P_{B} of entity B.

— A computes $\mathbf{c} = \mathbf{mG}' + \mathbf{e}$, where $\mathbf{e}$ is random binary vector of length n and weight t.

3. Decryption.

Entity B decrypts $\mathbf{c}$, using his private key S_{B}.

— Compute $\mathbf{x} = \mathbf{cP}^{-1}$.

— Compute $\psi(\mathbf{x}) = \mathbf{m}$.

McEliece proposed to use Goppa codes as private code. This codes are cyclic and can be decoded in polynomial time by decoders constructed using algebra for polynomials [4, 5]. Public key here is the code equivalent to private code (i.e. obtained by the coordinates permutation). It is supposed that the code equivalent to Goppa code can not be distinguished from the random code, though it is known that this is not true in some cases [7]. Additional requirement to private code is that code construction should allow exponentially large key space for given parameters of the code.

Analysis of McEliece cryptosystem may be performed in two directions. First, this is the recovering of the private code's structure from the public code. In fact this is the analysis of masking transformation, which is permutation in case of McEliece cryptosystem. In worst case this requires considering all permutations of length n, which is clearly infeasible.

Second, and this is counted as the main attack on McEliece system, is an attempt to correct t errors in ciphertext $\mathbf{c}$ and find the codeword in code $\mathbf{G}'$, i.e. solving the decoding problem in the sphere of radius t for the code which considered as random. Best known approach to solve this task for today is information set decoding [8–11]. Note that equivalent code has the same minimal distance as initial code, so bounded-distance decoding will find the correct codeword with probability 1, so the attack is limited only by computational complexity.

In the first variant of the system McEliece proposed to use (1024, 524)-code correcting 50 errors. Comparatively up-to-date review of decoding methods given in [8] mentioned that this parameters are correspondent to cryptocomplexity equal to 2^{53}, to achieve level of 2^{94} the matrix size should be 1036×2048 (correcting 92 errors), and matrix size 2056×4096 (correcting 170 errors) provide system strength of 2^{171}. In general the key sizes of this system have the order of hundreds of thousands bits. In many situations this is not excessive requirement, but traditional point of view is that this is the main drawback of the McEliece system.

The following directions of McEliece system improvement may be formulated:

1) reducing the key sizes by usage of special classes of Goppa codes, or alternative error-correction codes;

2) increasing of system's strength, first of all by strengthening the masking transformation between public and private keys.

In 1986 H. Niederreiter proposed the code-based system, for which later its equivalence to McEliece system was proven [8], but having some practical advantages. In this paper we do not consider this approach.

In the last decade the significant direction of McEliece system evolution is usage of block-circulant matrices for public and private codes, such matrices define the so called quasi-cyclic (QC) codes and allow significant reduction of key sizes during storage and transportation by means of circulant structure. To provide the polynomial-time decoding procedure, the private key is selected as sparse matrix, in this keys the decoding algorithms for low-density parity-check (LDPC) codes may be used [8, 12, 13]. In some cases of such systems the masking transformation is no longer the permutation matrix and selected in a special way (however, this transformation matrices should also be sparse to avoid large increasing of the number of errors corrected during decryption), however, in all such systems the underlying problem is bounded-distance decoding.

Public-key cryprosystems based on complete decoding problem

As it was mentioned in the previous section, there are modifications of McEliece system considering other classes of codes, and in some cases the special transformation matrices are considered instead of permutation matrix to hide the secret key in the public key. However, the fundamentally qualitative modification would be consideration totally random matrix for masking operation instead of permutation matrix or its analogues. In this case not only the public key is no longer defines the equivalent codes, but the number of qualitative new properties of the system are appeared.

Initially this approach was proposed by E. Krouk in 1993 [14] and later considered in the number of publications [15]. Let us describe the general structure of the system.

1. Key generation.

Each entity U performs the following.

— Select generator $(k \times n)$-matrix $\mathbf{G}$ of linear codes, for which the polynomial-time decoding procedure ψ is known, which corrects errors from the set E.

— Compute $\mathbf{G}' = \mathbf{GM}$, где $\mathbf{M}$ — $(n \times n)$ non-singular matrix.

— Define the set $E' = \{\mathbf{e}' \colon \mathbf{e}' = \mathbf{eM},\ \mathbf{e} \in E\}$.

— Public key is $P_{\mathrm{U}} = (\mathbf{G}', E')$, private key is $S_{\mathrm{U}} = (\mathbf{G}, \mathbf{M})$.

2. Encryption.

Entity A encrypts k-bit message $\mathbf{m}$, using authentic public key P_{B} of entity B.

— A computes $\mathbf{c} = \mathbf{mG}' + \mathbf{e}'$, where $\mathbf{e}' \in E'$.

3. Decryption.

Entity B decrypts $\mathbf{c}$, using his private key S_{B}.

— Compute $\mathbf{x} = \mathbf{cM}^{-1}$.

— Compute $\psi(\mathbf{x}) = \mathbf{m}$.

The problem with implementation of described system is that there are two generalized sets of errors: the set E of errors, which should be corrected during decryption and the set E' of errors used during encryption (in McEliece system both sets consist of error vectors of weight t). Both sets should be exponentially large to avoid brute force, and at the same time they should have compact representation.

In described variant the vectors from these sets are connected with help of multiplication by $\mathbf{M}$, but this matrix is the part of private key, while vectors from E' should be generated by the party possessing only the public key.

From the other hand, suppose that this problem is somehow solved. Then, if we consider for example that E is the set of vectors of weight t, as in McEliece system, vector $\mathbf{e}' = \mathbf{eM}$, where $\mathbf{M}$ is random, has random weight, which is more probable close to $n/2$. Besides, the matrix $\mathbf{G}' = \mathbf{GM}$ defines the code with minimal distance which is more probable less than in private code $\mathbf{G}$. Thus, if the system is analyzed through decoding (to reconstruct $\mathbf{m}$ from $\mathbf{c}$), one should correct approximately $n/2$ errors in the code with probably small minimal distance, instead of solving the problem (2). This may lead to the situation when error vector is not within coset leaders at all, thus even solving the minimum distance decoding problem (1), which is complete decoding problem, will not give the correct codeword. In this case the problem of breaking the system is at least not simpler than complete decoding (though one should take in mind the possibility of breaking the system through analysis of the structure of public codes and matrix $\mathbf{M}$), thus we will call such system as based on the problem of complete decoding.

Next we describe more practical variant of the system based on the problem of complete decoding, giving an example of defining E and E' [15].

1. Key generation.

Each entity U performs the following.

— Select $(k \times n)$-generator matrix $\mathbf{G}$ of linear (n, k)-code, for which polynomial-time algorithm for correcting errors from some set E is known.

— Select random non-singular $(n \times n)$-matrix $\mathbf{M}_2$.

— Define the set $\tilde{E}$ and matrix $\mathbf{M}_1$ such that for any $\tilde{\mathbf{e}} \in \tilde{E}$ the vector $\tilde{\mathbf{e}}\mathbf{M}_1$ belongs to E (note that $\mathbf{M}_1$ may be singular).

— Compute $\mathbf{M} = \mathbf{M}_1\mathbf{M}_2$ (note that $\mathbf{M}$ singular if $\mathbf{M}_1$ is singular).

— Compute $(k \times n)$-matrix $\mathbf{G}' = \mathbf{G}\mathbf{M}_2$.

— Public key is $P_{\text{U}} = (\mathbf{G}', \mathbf{M}, \tilde{E})$, private key is $S_{\text{U}} = (\mathbf{G}, \mathbf{M}_1, \mathbf{M}_2)$.

2. Encryption.

Entity A encrypts k-bit message $\mathbf{m}$, using authentic public key P_{B} of entity B.

— A computes $\mathbf{c} = \mathbf{m}\mathbf{G}' + \mathbf{e}'$, where $\mathbf{e}' = \tilde{\mathbf{e}}\mathbf{M}$ for random $\tilde{\mathbf{e}} \in \tilde{E}$.

3. Decryption.

Entity B decrypts $\mathbf{c}$, using his private key S_{B}.

— Compute $\mathbf{x} = \mathbf{c}\mathbf{M}_2^{-1}$.

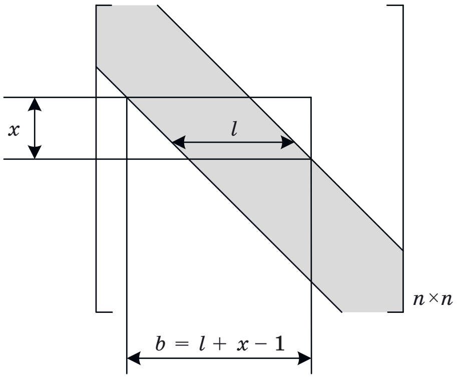
— Compute $\psi(\mathbf{x}) = \mathbf{m}$.

In this variant of the system the set E' is defined by vectors $\tilde{\mathbf{e}}\mathbf{M}$, which in turn requires effective description of $\tilde{E}$. Besides, the matrix $\mathbf{M}_1$ should be determined, mapping vectors from $\tilde{E}$ into E.

In particular, the set E itself may be selected as $\tilde{E}$, for example, consisting of all vectors of fixed weight, as in classical McEliece system. In this paper we consider the variant of the system which based on error-correcting codes which correct error bursts.

The effect of grouping errors in bursts (or packets) is typical for the most real communication channels, however, the codes that can correct such erroneous combinations are less investigated, and in practice the data transmitted via the channel is decorrelated using the interleaving procedure, and then the codes for independent errors correction are applied. In the case of cryptosystem development the errors in bursts may be formed artificially, in this case the positions and lengths of the bursts may be controlled.

The term of error burst itself may be defined in different ways. In this paper we define the burst of length b as binary error vector $\mathbf{e} = (e_0, \ldots, e_{n-1})$, in which the last non-zero element is placed no more than in b positions from the first. That is, if i is the minimal index for which $e_i = 1$, and j is maximal such index, then $\mathbf{e}$ forms the (single) error burst of length $b = j - i + 1$ at position i (thus two adjacent non-zero element form the packet of length 2). We will assume that positions of $\mathbf{e}$ from index i to j are filled by 1 and 0 with probability 1/2. Note that under the term “burst” one may consider not only the overall sequence $\mathbf{e}$, but its erroneous subsequence $(e_i, \ldots, e_j)$ without leading and ending zeros, the concrete sense of this term will be clear from the context.



■ Matrix $\mathbf{M}_1$

Similar to the fact that the minimal distance d_0 defines the maximal number t of independent errors, which can be corrected in any combination by minimal distance decoding, for each linear code the maximal correctable burst length b may be determined — this means that all possible error bursts of length no more than b are in different cosets and may be chosen as leaders. However, as it was mentioned earlier, finding the minimal distance of the random code is NP-hard, while maximal correctable burst length may be found in polynomial-time, using procedure from [16] (though the degree of the polynomial is rather large).

Let the set E consists of vectors which form error bursts of length no more than b. As the set $\tilde{E}$ we will also consider the set of bursts, but their length may differ from b and is defined by $\mathbf{M}_1$.

Consider $\mathbf{M}_1$ as $(n \times n)$-matrix in Figure. Here positions filled by random binary digits are marked in grey, other positions are zero. Clearly, such matrix through multiplication by it defines the mapping from bursts of length x into bursts of length b.

Then the above system may be additionally determined as follows:

— the set E: set of error bursts of length no more than b;

— matrix $\mathbf{G}$ defines the code, for which the polynomial-time procedure of correcting the error bursts of length b is known;

— the set $\tilde{E}$: set of error bursts of length no more than x. Clearly, public key is $P_{\text{U}} = (\mathbf{G}', \mathbf{M}, x)$.

In the next sections we will consider the selection of the code for the proposed system and estimation of its parameters.

Selection of the code for the system

Estimation of the quantitative parameters of the system considered in the previous section: burst lengths x and b, cardinalities of sets E' and $\tilde{E}$, and finally selection of k and n, defining the key sizes, depends on the selection of class of burst-error-correction code. This class should contain exponential number of codes for given b, k, n, and admit polynomial-time procedure of correcting the bursts of length b. One of the variant of such a class is the class of low-density parity-check codes.

Low-density parity-check codes (LDPC codes) were proposed by R. Gallager in early 60-s [8, 12]. LDPC-code is defined by its parity-check matrix $\mathbf{H}$, containing low number of nonzero elements. The term "low number" is not formally defined, moreover, in the number of works on modified McEliece systems based on such codes the term "middle density" (MDPC) is used [7, 17, 18], but in both cases we may admit that we consider the codes with relatively sparse parity-check matrix, for which the decoders utilizing its sparseness show rather high correcting capability (low error probability). In general, LDPC codes are usually defined and analyzed as probabilistic ensembles of random codes with specific parameters, which is additional advantage for their usage as secret keys in code-based systems.

One of the most often used construction of LDPC codes is block-permutation construction, where the parity-check matrix has the form

$$\mathbf{H}=\begin{bmatrix}\mathbf{H}_{1,1} & \mathbf{H}_{1,2} & \dots & \mathbf{H}_{1,\rho}\\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} & \dots & \mathbf{H}_{2,\rho}\\ \dots & \dots & \dots & \dots\\ \mathbf{H}_{\gamma,1} & \mathbf{H}_{\gamma,2} & \dots & \mathbf{H}_{\gamma,\rho}\end{bmatrix},$$

where $\mathbf{H}_{i,j}$ are sub-blocks of some structure. Usually some degree of $(m \times m)$-matrix of cyclic permutation is used as sub-blocks:

$$\mathbf{C}=\begin{bmatrix}0 & 0 & 0 & \dots & 0 & 1\\ 1 & 0 & 0 & \dots & 0 & 0\\ 0 & 1 & 0 & \dots & 0 & 0\\ \dots & \dots & \dots & \dots & \dots & \dots\\ 0 & 0 & 0 & \dots & 1 & 0\end{bmatrix},$$

then the parity-check matrix of LDPC codes has the form

$$\mathbf{H}=\begin{bmatrix}\mathbf{C}^{i_{11}} & \mathbf{C}^{i_{12}} & \dots & \mathbf{C}^{i_{1\rho}}\\ \mathbf{C}^{i_{21}} & \mathbf{C}^{i_{22}} & \dots & \mathbf{C}^{i_{2\rho}}\\ \dots & \dots & \dots & \dots\\ \mathbf{C}^{i_{\gamma 1}} & \mathbf{C}^{i_{\gamma 2}} & \dots & \mathbf{C}^{i_{\gamma\rho}}\end{bmatrix}. \quad (3)$$

It is known that some specific combinations of non-zero elements in the parity-check matrix may degrade the LDPC code decoder's performance, the simplest restriction to avoid some "bad" combinations is absence of two rows or two columns in the parity-check matrix having more than one common non-zero positions. If this is holds and we consider the matrix as incidence matrix of bipartite graph (the so called Tanner graph) there are no cycles of length 4 in the graph (for simplicity we will say that there are no cycles of length 4 in the matrix).

Traditionally LDPC codes are used to correct independent errors, however, in [16, 19, 20] the capability of these codes (in particular, block-permutation constructions) to correct bursts of errors was analyzed. In [16] both the procedure of determining the maximal length of correctable burst and decoding procedure for block-permutation LDPC code are described.

For the system described in the previous section we will use the ensemble of codes defined by (3). The ensemble is defined by the values of γ, ρ and m. As the additional requirement we demand the absence of 4-cycles in matrix (3).

As follows, for the fixed values of γ, ρ and m, which are selected mainly from the cryptocomplexity point of view, the probability of random selection of matrix with no cycles of length 4 should be estimated, as well as expected correctable burst lengths.

The software for determining the correctable burst length was implemented in Microsoft Visual Studio Enterprise 2017 using C++, Release x64. Experiments were hold using the computer with Windows 10 Pro, 16 Gb RAM, CPU Intel Core i7-4770K@3,50 GHz.

The matrices are considered with 3×4, 4×6, 4×8 blocks, with block sizes $m = 20, 31, 50, 61, 110, 127$. The distribution of correctable burst lengths, as well as the average time of determining the burst length are given in Table 1. For each set of parameters 100 random matrices were generated without cycles of length 4. One may note that for the block sizes which are prime numbers the lengths of correctable burst in all experiments are $b = m - 1$ [this is the maximal possible length of correctable burst for the block-permutation codes with parity-check matrix (3) with block size m]. In other cases it was found that for considered values of γ and ρ the burst lengths correctable by random codes is not significantly less than block size m, i.e. $b \approx m$.

The estimation of probability P of selecting the matrix containing cycles of length 4 in the Tanner graph was also performed, the results are given in Table 2. As can be seen from the table, when the block sizes are small, the probability of selecting the matrix with cycles of length 4 is rather high. But this probability decreases with block sizes growth, besides, determination of cycle existence in all considered cases takes less than microsecond, thus even if the probability of matrix without cycles

■ ***Table 1.*** Examples of estimations of correctable burst lengths distributions and average processing time for different number of blocks in the parity-check matrix

$\gamma \times \rho$	Parameter	Block size m					
		20	31	50	61	110	127
3×4	Length b	19 — 10% 18 — 64% 16 — 19% 15 — 7%	30 — 100%	49 — 9% 48 — 75% 45 — 12% 40 — 4%	60 — 100%	109 — 8% 108 — 70% 105 — 14% 100 — 6% 99 — 2%	127 — 100%
	Time, s	0.01	0.04	0.38	0.63	11.64	12.43
4×6	Length b	19 — 21% 18 — 71% 16 — 4% 15 — 4%	30 — 100%	49 — 13% 48 — 80% 45 — 7%	60 — 100%	109 — 11% 108 — 81% 105 — 7% 100 — 1%	126 — 100%
	Time, *s*	0.04	0.18	2.68	2.86	50.92	80
4×8	Length b	19 — 2% 18 — 83% 16 — 12% 15 — 3%	30 — 100%	49 — 1% 48 — 88% 45 — 11%	60 — 100%	109 — 1% 108 — 84% 105 — 12% 100 — 3%	126 — 100%
	Time, s	0.10	0.44	6.26	6.75	80	171

■ ***Table 2.*** Probability of existence of cycle of length 4

$\gamma \times \rho$	Parameter	Block size m					
		20	31	50	61	110	127
3×4	Probability P	0.62	0.46	0.31	0.26	0.15	0.13
	Time, µs	0.25	0.28	0.23	0.26	0.26	0.28
4×6	Probability P	0.99	0.96	0.85	0.78	0.57	0.51
	Time, µs	0.55	0.58	0.61	0.52	0.57	0.59
4×8	Probability P	0.999	0.997	0.97	0.95	0.8	0.75
	Time, µs	0.58	0.85	0.93	0.71	0.83	0.8

of length 4 is rather small, this matrix may be generated expectably fast by the sequence of random guesses (total number of such matrices is exponentially high). For example, for $\gamma = 4$, $\rho = 8$, $m = 20$ the probability of absence of 4-cycles is less than 0.001, but total number of matrices is $20^{32} \approx 2^{138}$, and appropriate random matrix is easy to find.

Summarizing, the results show that for small values of γ and ρ it takes not a lot of time to find the random matrix (3) without 4-cycles and with $b = m - 1$.

Estimation of system's parameters

In this section we estimate the parameters of the cryptosystem, basing on required security level. We will consider the following attacks, which complexity should be exponential of order not less than 2^{128}:

— search on private matrices;
— search within the set E';
— search within the set $\tilde{E}$.

Let us consider the search on private matrices. There are $m^{\gamma\rho}$ matrices (3) of $\gamma \times \rho$ blocks, where each block is defined by integer from the set $\{0, ..., m - 1\}$. For example we will take matrices with 3×6 and 4×8 blocks. From $m^{18} = 2^{128}$ and $m^{32} = 2^{128}$ we have the correspondent block sizes $m \approx 2^7 = 128$ and $m = 2^4 = 16$. To increase the probability of selecting the code with maximal correctable burst length we set m as prime number, i.e. $m = 127$ and $m = 17$. Particular selection of block size should be additionally agreed with the length b of correctable burst, which should be provided, so given estimations may be considered as lowest possible values for m.

Let us now consider brute-force search within the set E'. It consists of vectors $\mathbf{e}' = \tilde{\mathbf{e}}\mathbf{M}$, where $\mathbf{M} = \mathbf{M}_1\mathbf{M}_2$. Despite the special construction of matrix $\mathbf{M}_1$ (see Figure), $\mathbf{M}_2$ is random matrix, thus even though $\tilde{\mathbf{e}}$ is error burst of length x, $\mathbf{e}'$ is random vector with expected weight $n/2$, which makes impossible both enumerating these vectors and breaking the system by decoding — any code including $\mathbf{G}'$ is not able to correct error vectors of such weight.

Finally, consider the set $\tilde{E}$. It consists of vectors which are bursts of length x. If starting position of the burst is fixed, there are about 2^x such vectors (in fact we should fill the positions within burst by random bits with probability 1/2 and obtain the bursts of weight near, so the number of such vectors is slightly less than 2^x). The number of burst locations (starting positions) within the error vector $\tilde{\mathbf{e}}$ is $n - x + 1$. This number is not very large, but we take it into account. Thus the complexity of brute force search of $\tilde{\mathbf{e}}$ is given by $(n - x + 1)2^x = 2^{128}$ (this equation is approximate, we do not take into account the weight of bursts, starting and ending 1's and so on).

Consider example with $\gamma = 3$, $\rho = 6$. Select $m = 127$, then the correctable burst length is $b = 126 = l + x - 1$, where l is the width of diagonal in matrix $\mathbf{M}_1$ in Figure. For such parameters we have $n = m\rho = 762$, and $(763 - x)2^x = 2^{128}$ and $x \approx 119$, hence $l = 8$.

For such parameters $\mathbf{M}_1$ is matrix 762×762, containing diagonal of width $l = 8$. Analysis of how the structure of $\mathbf{M}_1$ affects the system's strength is important question, but it does not considered in this paper since it requires more thorough and sophisticated analysis. Nevertheless, it seems that usage of rather large matrix with relatively thin diagonal of random elements while other elements are zero may be not secure. Thus consider the possibility of increasing the width of diagonal in $\mathbf{M}_1$.

Let us take, for example, $l = 30$. Then, having $x = 119$, the length of the burst correctable by the secret code should be $b = 149$ and hence the block size $m \ge 150$. We obtain the following parameters: $\gamma = 3$, $\rho = 6$, $l = 30$, $b = 149$, $m = 150$, $n = 900$, $x = 119$, then the number of bursts of length x is estimated as $2^{128.6}$, which corresponds to the required security level. To define the matrix $\mathbf{H}$ (3) it is enough to store only degrees of correspondent matrices $\mathbf{C}$, for our parameters we get $3 \cdot 6 \cdot \lceil \log_2 150 \rceil = 144$ bits. The size of $\mathbf{G}'$ (public key, which is more important from the point of view of key size), is 450×900 bits (which is less than initial parameters of McEliece system with key size 450×900 and cryptocomplexity 2^{53}). Note that we should also count the size of matrix $\mathbf{M}$ since it's the part of public key.

Similarly consider the case $\gamma = 4$, $\rho = 8$, with $m = 17$, which was defined earlier, and $b = 16$, $n = 136$. Then $x \approx 125$, and taking into account $b = l + x - 1$ it is impossible to have $b = 16$. Hence, the block size m should be significantly increased, as well as the burst length b.

Set $x = 125$ and $l = 30$, then we have $b = 155$ and $m \ge 156$, this gives parameters: $\gamma = 4$, $\rho = 8$, $l = 30$, $b = 155$, $m = 156$, $n = 1248$, $x = 125$, then the number of bursts of length x is 2^{135}. The size of private key is 256, public matrix $\mathbf{G}'$ — 624×1248 bit.

As can be seen from the estimations, the cryptocomplexity of proposed system depends on the number of bursts of length x, which defines the complexity of brute force search within the set $\tilde{E}$, and also on possibility of attacking the structure of $\mathbf{M}_1$, which depends on the value of l. These parameters define the values of b and m, such that enumerating the matrices $\mathbf{H}$ should be infeasible.

In Table 3 the key sizes, number of errors that should be decoded by adversary to break the system, cryptocomplexity of McEliece system, system based on quasi-cyclic LDPC codes (QC-LDPC) described in [8], and proposed system are collected.

Note that the public key size is defined by the size of $(k \times n)$-matrix $\mathbf{G}'$, however in proposed system the $(n \times n)$-matrix $\mathbf{M}$ is also the part of the public key, at the same time in QC-LDPC system significantly larger codes are used, but using the block-circulant structure of the matrix (including public matrix) the required storage for the keys may be significantly reduced. From the other hand, one should make the distinction between storage needs and memory which is used during encryption and especially decryption processes, when the decoding procedure should be

■ ***Table 3.*** Comparison of code-based cryptosystems

System	k	n	t	Complexity	Attack	Public key size, Kbyte	Private key size, Kbyte
McEliece	524	1024	50	2^{53}	Bounded-distance decoding (t errors)	67	67
	1036	2048	92	2^{94}		265	265
	2056	4096	170	2^{171}		1052	1052
QC-LDPC	9857	19 714	134	2^{128}		1.2	1.2
Bursts (proposed)	450	900	450	2^{128}	Complete decoding (n/2 errors) or brute force search on $\tilde{E}$	152	0.144
	624	1248	624	2^{135}		291	0.25

used. With increasing of matric sizes the required memory is also increases, but estimations depend on particular implementations and their optimizations.

At the same time, further development of the proposed system in the direction of using quasi-cyclic codes is of interest.

Conclusion

In the paper the code-based cryptosystem is proposed which uses burst-correction codes. The underlying hard mathematical problem is complete decoding problem. It is supposed that systems based on this problem can achieve better cryptocomplexity than code-based McEliece cryptosystem based on bounded-distance decoding.

The future investigations and development can be made in using quasi-cyclic codes and considering the version of the system in Niederreiter mode.

Financial support

The paper was prepared with the financial support of the Russian Science Foundation, project No. 22-19-00305 “Spatial-temporal stochastic models of wireless networks with a large number of users”.

References

1. Stinson D. R., Paterson M. B. *Cryptography. Theory and Practice*. CRC Press, 2018. 598 p.
2. Cormen T. H., Leiserson C. E., Rivest R. L., Stein C. *Introduction to Algorithms*. MIT Press, 2022. 1312 p.
3. *NIST Post-quantum Cryptography Project*. Available at: https://csrc.nist.gov/projects/post-quantum-cryptography (accessed 21 April 2022).
4. Lin S. *Fundamentals of Classical and Modern Error-Correcting Codes*. Cambridge University Press, 2022. 800 p.
5. Moon T. K. *Error Correction Coding: Mathematical Methods and Algorithms*. Wiley, 2020. 992 p.
6. Chailloux A., Debris-Alazard T., Etinski S. Classical and quantum algorithms for generic syndrome decoding problems and applications to the lee metric. *Proc. 12th Intern. Conf. “Post-Quantum Cryptography”, PQCrypto 2021*, J. H. Cheon, J.-P. Tillich (eds), LNCS, 2021, vol. 12841, pp. 44–62. doi: 10.1007/978-3-030-81293-5_3
7. Sendrier N. Code-based cryptography: State of the art and perspectives. *IEEE Security & Privacy*, 2017, vol. 15, no. 4, pp. 44–50. doi: 10.1109/MSP.2017.3151345
8. Baldi M. *QC-LDPC Code-Based Cryptography*. Springer, 2014. 120 p.
9. Canto Torres R., Sendrier N. Analysis of information set decoding for a sub-linear error weight. *Proc. 7th Intern. Conf. “Post-Quantum Cryptography”, PQCrypto 2016*, T. Takagi (ed), LNCS, 2016, vol. 9606, pp. 144–161. doi:10.1007/978-3-319-29360-8_10
10. Both L., May A. Decoding linear codes with high error rate and its impact for LPN security. *Proc. 9th Intern. Conf. “Post-Quantum Cryptography”, PQCrypto 2018*, T. Lange, R. Steinwandt (eds), LNCS, 2018, vol. 10786, pp. 25–46. doi:0.1007/978-3-319-79063-3_2
11. Kirshanova E. Improved quantum information set decoding. *Proc. 9th Intern. Conf. “Post-Quantum Cryptography”, PQCrypto 2018*, T. Lange, R. Steinwandt (eds), LNCS, 2018, vol. 10786, pp. 507–527. doi:10.1007/978-3-319-79063-3_24
12. Lin S., Ryan W. *Channel Codes: Classical and Modern*. Cambridge University Press, 2009. 710 p.
13. Baldi M., Barenghi A., Chiaraluce F., Pelosi G., Santini P. LEDAkem: A post-quantum key encapsulation mechanism based on QC-LDPC codes. *Proc. 9th Intern. Conf. “Post-Quantum Cryptography”, PQCrypto 2018*, T. Lange, R. Steinwandt (eds), LNCS, 2018, vol. 10786, pp. 3–24. doi:10.1007/978-3-319-79063-3_1
14. Krouk E. A new public-key cryptosystem. *Sixth Joint Swedish-Russian Intern. Workshop on Information Theory*, Moelle, Sweden, 1993, pp. 285–286.
15. Krouk E., Ovchinnikov A. Code-based public-key cryptosystem based on bursts-correcting codes. *The Thirteen Advanced Intern. Conf. on Telecommunications*, 2017, pp. 93–95.
16. Veresova A. M., Ovchinnikov A. A. About one algorithm for correcting bursts using block-permutation LDPC-codes. *2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, Saint-Petersburg, Russia, 2019, pp. 1–4.
17. Sendrier N., Vasseur V. About low DFR for QC-MDPC decoding. *Proc. 11th Intern. Conf. “Post-Quantum Cryptography”, PQCrypto 2020*, J. Ding, J.-P. Tillich (eds), LNCS, 2020, vol. 12100, pp. 20–34. doi:10.1007/978-3-030-44223-1_2
18. Eaton E., Lequesne M., Parent A., Sendrier N. QC-MDPC: A timing attack and a CCA2 KEM. *Proc. 9th Intern. Conf. “Post-Quantum Cryptography”, PQCrypto 2018*, T. Lange, R. Steinwandt (eds), LNCS, 2018, vol. 10786, pp. 47–76. doi:10.1007/978-3-319-79063-3_3
19. Krouk E. A., Ovchinnikov A. A. Exact burst-correction capability of gilbert codes. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2016, no. 1, pp. 80–87 (In Russian). doi:10.15217/issn1684-8853.2016.1.80
20. Ovchinnikov A., Fominykh A. About burst decoding for block-permutation LDPC codes. *Internet of Things, Smart Spaces, and Next Generation Networks and Systems: 20-th Intern. Conf., NEW2AN 2020, and 13-th Conf., ruSMART 2020*, Saint-Petersburg, Russia, 2020, pp. 393–401. doi:10.1007/978-3-030-65726-0_35

УДК 003.26
doi:10.31799/1684-8853-2022-3-45-54

Вариант постквантовой системы на основе кодов, исправляющих пакеты ошибок, и задачи полного декодирования

А. А. Овчинников[a], канд. техн. наук, доцент, orcid.org/0000-0002-8523-9429, mldoc@guap.ru
[a]Санкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ

Введение: важным направлением в исследовании постквантовых систем, устойчивых к квантовым вычислениям, является кодовая криптография на основе задач теории помехоустойчивого кодирования. Улучшение существующих кодовых систем может вестись как в практической части (уменьшение размеров ключей), так и с точки зрения использования более трудных математических кодовых задач. **Цель:** построение кодовой системы с открытым ключом на основе низкоплотностных кодов, исправляющих пакеты ошибок; оценка параметров полученной системы. **Результаты:** предложен вариант кодовой системы на основе случайных блочно-перестановочных низкоплотностных кодов. Стойкость системы предполагается основанной на задаче полного декодирования, что является более сложной математической задачей по сравнению с существующими системами. При этом с высокой вероятностью анализ системы на основе методов декодирования вообще не представляется возможным, что как повышает перспективную стойкость системы, так и позволяет уменьшить размеры ключей. Проведена оценка выбора кодов с требуемыми характеристиками, рассматриваются подходы к выбору параметров предложенной системы на основе требуемого уровня стойкости. **Практическая значимость:** предложенная система позволяет уменьшить размеры открытых ключей по сравнению с классической системой МакЭлиса при сравнимой стойкости, при этом используемая трудная математическая задача представляется более устойчивой к перспективным атакам.

Ключевые слова — постквантовая криптография, кодовые системы, коды с малой плотностью проверок на четность, исправление пакетов ошибок.

Для цитирования: Ovchinnikov A. A. The variant of post-quantum cryptosystem based on burst-correcting codes and on the complete decoding problem. *Информационно-управляющие системы*, 2022, № 3, с. 45–54. doi:10.31799/1684-8853-2022-3-45-54

For citation: Ovchinnikov A. A. The variant of post-quantum cryptosystem based on burst-correcting codes and on the complete decoding problem. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2022, no. 3, pp. 45–54. doi:10.31799/1684-8853-2022-3-45-54

УВАЖАЕМЫЕ АВТОРЫ!

Научная электронная библиотека (НЭБ) продолжает работу по реализации проекта SCIENCE INDEX. После того как Вы зарегистрируетесь на сайте НЭБ (http://elibrary.ru/defaultx.asp), будет создана Ваша личная страничка, содержание которой составят не только Ваши персональные данные, но и перечень всех Ваших печатных трудов, имеющихся в базе данных НЭБ, включая диссертации, патенты и тезисы к конференциям, а также сравнительные индексы цитирования: РИНЦ (Российский индекс научного цитирования), h (индекс Хирша) от Web of Science и h от Scopus. После создания базового варианта Вашей персональной страницы Вы получите код доступа, который позволит Вам редактировать информацию, помогая создавать максимально объективную картину Вашей научной активности и цитирования Ваших трудов.