НАУЧНЫЙ ОТДЕЛ

ИНФОРМАТИКА

A Method of Protected Distribution of Data Among Unreliable and Untrusted Nodes

Yu. V. Kosolapov, F. S. Pevnev

Yury V. Kosolapov, https://orcid.org/0000-0002-1491-524X, Institute of Mathematics, Mechanics, and Computer Science named after of I. I. Vorovich, Southern Federal University, 8a Milchakova St., Rostov-on-Don 344090, Russia, itaim@mail.ru

Fedor S. Pevnev, Institute of Mathematics, Mechanics, and Computer Science named after of I. I. Vorovich, Southern Federal University, 8a Milchakova St., Rostov-on-Don 344090, Russia, fes_21@mail.ru

We consider a model of protecting the confidentiality and recoverability of data in a distributed storage system. It is assumed that informational blocks are coded into the code blocks. Then the blocks are divided into parts and distributed among repositories of the distributed storage. A modification of the code noising method is constructed which simultaneously provides computational resistance to coalition attacks on confidentiality of stored data. Moreover, the modification also provides protection from the failure of a part of the storage nodes. Confidentiality protection is provided for coalitions of greater cardinality than in the case of using the classical method of code noising. It is shown that computational resistance is based on the complexity of solving one well-known problem of theoretical coding.

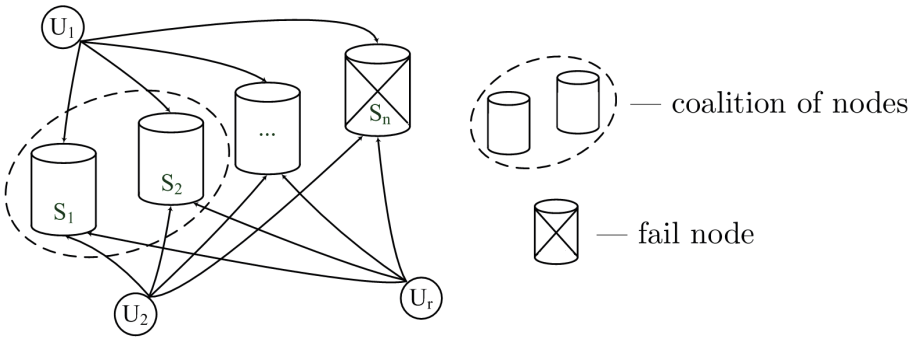
Keywords: wiretap channel, distributed secure storage, coalition attack.

Received: 05.10.2018 / Accepted: 21.05.2019 / Published: 31.08.2019

DOI: https://doi.org/10.18500/1816-9791-2019-19-3-326-337

INTRODUCTION

Let us consider a model of safe data storage on n independent and in general untrusted repositories $S_1, \ldots, S_n$ (Figure). Further these repositories are sometimes referred to as nodes. We consider cloud repositories like Google Drive,



The distributed storage system

Yandex.Disk, etc. to be such independent storages. The users are able to write their data into each of n repositories and read data from at least $\nu(\in \mathbb{N})$ ones (inaccessible repositories are crossed out on the Figure). We assume that adversary coalition contains no more than $\mu(\in \mathbb{N})$ repositories (referred to as participants of coalition) and are able to obtain data from each of them (coalition is marked with a dashed line on the Figure). The parameters n, ν and μ are known to everybody. The challenge for the developers of a protection system is choosing the transformation of protected data before distributing it among repositories. On the one hand, this transformation should provide confidentiality of protected data against coalition of cardinality μ or less, on the other hand, it should provide a possibility of recovering the protected data when any $n - \nu$ repositories are inaccessible. The coding method is considered to be not secret. We are interested in non-cryptographic methods, because in this case it is not necessary to support the life cycle of cryptographic keys.

The store model described above is actually the research subject of [1]. In [1] transformation of protected data is a *code noising* method (in terminology [2]) based on a pair of linear codes $(\widetilde{C}, C)$ where $[n, l, \widetilde{d}]$-code $\widetilde{C}$ with length n, dimension l, and code distance $\widetilde{d}$ contains $[n, l-k, d]$-code C, $k < l$. In [1] both codes are MDS-codes (Maximum Distance Separable codes). Code noising method is optimal for this store model if $n - \nu \leqslant \widetilde{d} - 1$ and

$$\mu \leqslant l - k \tag{1}$$

(see results in [3–6]). In this case the confidentiality is provided in theoretical-informational sense if protected data is uniformly random distributed. The pairs of MDS-codes are also optimal if availability of the data storage is limited [7], or if coalition has an access to an additional part of protected data [8]. Some experimental estimations of code noising resistance in distributed storage are explored in [9], but the observer has identified an attack algorithm in that case.

The article [10] considers a *repetitive interception attack* against the classical code noising method. It is assumed that the observer has the opportunity to notice several partial code blocks corresponding to *one* unknown informational block. In the article [10] it is also assumed that different code blocks are observed on different subsets of coordinates. The repetitive interception attack is successful if condition (1) is wrong [10]. Thus, in the distributed storage model the coalition of repositories is able to attack confidentiality effectively with the repetitive interception attack. This attack is possible if the system similar to one described in [1] is used and the condition (1) is wrong. We offer a modification of the code noising method which provides high resistance to repetitive interception even in case when condition (1) is wrong.

Our solution is based on the regular change of the coding map. Synchronization of the sender and the receiver is not required, however, the sender needs to additionally send the information about the mapping used. We use an approach usual in cryptography to estimate the resistance of proposed method. According to this method it is enough to reduce the task by breaking it into several (usually well-known) matematical problems. In the present paper the resistance of the constructed method is based on the complexity of one theoretical coding problem.

The article consists of introduction, two sections, and conclusion. The first section describes an analytical model (data storage model), a code noising method and its modification. The second section analyzes the application of the constructed modified method in a distributed storage system. An estimate is obtained for the number of storages that may fail without affecting the possibility of correct recovering of informational blocks from uninjured repositories (nodes). Also the resistance of the modified method is analyzed for coalition attack.

1. DATA PROTECTION SCHEME

1.1. Analytical model

Let us briefly describe the data storage model proposed in [1]. Let $i \in \{1, ..., r\}$ and data from i-th source U_i be represented as informational blocks of k characters over a finite field $\mathbb{F}_q$. Each informational block is encoded independently into the code block of n characters from $\mathbb{F}_q$ via encoder Enc. Then all n symbols of code block $\mathbf{c} = (c_1, ..., c_n) \in \mathbb{F}_q^n$ are distributed in n repositories so that j-th symbol c_j is written to the repository with number $\pi_i(j)$ (or equivalently to the node $S_{\pi_i(j)}$) where

$$\pi_i : \{1, ..., n\} \to \{1, ..., n\} \tag{2}$$

is a permutation. The users independently choose permutations (2) which are not private. We also assume that the users know the permutation π_i while they obtain information from repositories. The permutations may appear different from block to block or from file to file or other way.

To extract one informational block the user reads characters of the corresponding code block from the repositories and then puts the whole block into the decoder Dec. The value of $\pi_i^{-1}(j)$-th coordinate is unknown for the user if the repository S_j is inaccessible (e.g. due to failure or injury). In this case we consider this coordinate to be erased and write symbol $*$ instead of its value. We assume that no more than $n-\nu$ repositories may be inaccessible while the user is reading data. As the repositories are supposed to be untrusted, we consider every node to be an *eavesdropper* which knows a value of only one coordinate in every code block. Other coordinates are considered to be erased. The participants of coalition of μ repositories will know values of μ coordinates in every code block. This set of coordinates may be different from block to block because of different permutations conducted (2). Therefore, the coalition has the opportunity to launch a repetitive interception attack from [10] if the classical code noising method is used.

1.2. Classical code noising method

The code noising method is used in [1] for keeping data safety against adversary coalition and inaccessibility of the repositories at the same time. We can describe this method in the following way. Let $\widetilde{C}$ be a linear $[n, l, \widetilde{d}]$-code with length n, dimen-

sion l, and code distance $\widetilde{d}$, $\widehat{C}$ and C are $[n,k]$-code and $[n,l-k]$-code respectively, $C \cap \widehat{C} = \mathbf{0} = (0,...,0)(\in \mathbb{F}_q^n)$ and direct sum $C \oplus \widehat{C}$ is equal to $\widetilde{C}$. Let $\widehat{G}$ and G be generating matrices of codes $\widehat{C}$ and C respectively. Code noising is the function $f : \mathbb{F}_q^k \times \mathbb{F}_q^{l-k} \to \widetilde{C}$,

$$f(\mathbf{m}, \mathbf{r}) = \mathbf{m}\widehat{G} + \mathbf{r}G = \mathbf{c}$$

where $\mathbf{m}(\in \mathbb{F}_q^k)$ is an informational block, $\mathbf{r}$ is vector which is chosen randomly and equiprobably from $\mathbb{F}_q^{l-k}$. Let

$$\mathrm{Dec}_{\widetilde{C}} : (\mathbb{F}_q \cup \{*\})^n \to \mathbb{F}_q^l$$

be a decoder which is able to correct no more than $\widetilde{d} - 1$ erasures in every code block and has vectors from $\mathbb{F}_q^l$ as output. One can try to obtain the informational block from the block $\mathbf{c}' \in (\mathbb{F}_q \cup \{*\})^n$ by applying the decoder $\mathrm{Dec}_{\widetilde{C}}$ to the $\mathbf{c}'$ and cutting off the last $l-k$ symbols of the decoder output.

Let us assume that every informational block $\mathbf{m}(\in \mathbb{F}_q^k)$ has an equal probability $p_M(\mathbf{m}) = 1/q^k$, i.e. random variable M is uniformly distributed over $\mathbb{F}_q^k$. As we can see in [1,4–6,8] the resistance of the code noising method strongly depends on pair $(\widetilde{C}, C)$. In fact for every pair $(\widetilde{C}, C)$ there exists a threshold $\mu_0(\in \mathbb{N})$ such that if the coalition (or eavesdropper) knows the values of no more than μ_0 coordinates of the code block it will not obtain any information about encoded informational block. Otherwise if $\mu > \mu_0$ the eavesdropper can get non-zero information. In this case there is at least one set of observed coordinates τ ($|\tau| = \mu$) which does not provide the whole set of informational blocks as candidates to be original informational block, i.e. the size of the provided set of candidates is less than q^k. So the eavesdropper may attempt to use repetitive interception attack from [10]. For example, as it is shown in [1], if $(\widetilde{C}, C)$ is a pair of MDS-codes then $\mu_0 = l - k$ (see (1)) where l is rank of $\widetilde{C}$ and $n-k$ is rank of C. The eavesdropper can easily recover the informational block knowing few partially erased code blocks if $\mu > \mu_0$ (see [10]). We propose a modification of the code noising method for counteracting a repetitive interception attack. We describe this modification in the next subsection. The defense ability will be described in subsection 2.

1.3. Modified code noising method

The main idea of the modified code noising method is periodic change of encoding functions in such way that the legal receiver can determine the exact encoding function using the received code blocks. Note that further the user is called a *legal receiver* if he or she has a permission to read data from the storage. At the same time an illegal eavesdropper cannot determine the exact function. Note that the idea of changing encoding functions is not new. The authors of [11] have used this idea creating the XtX encoding construction. They have assumed that the eavesdropper is able to obtain full data with errors (not erasures) and have analyzed properties of this construction such as code rate and security. The principal distinction of our scheme is using only one operation for providing security instead of two operations as in XtX construction.

We denote the set of numbers $\{1,...,n\}$ as $\underline{n}$. Let the set $\mathrm{supp}(\mathbf{a}) = \{i : a_i \neq 0\}$ be a support of vector $\mathbf{a} = (a_1,...,a_n)$ and the number $w(\mathbf{a}) = |\mathrm{supp}(\mathbf{a})|$ be a weight of this vector. For positive integers $n' \leqslant n$ the operator

$$\Pi_\tau : \mathbb{F}_q^n \to \mathbb{F}_q^{n'}$$

will be used as a projection operator on the set $\tau(\subseteq \underline{n})$. If $\tau = \{i_1, ..., i_{n'}\}$ and $\mathbf{a} = (a_1, ..., a_n)$, then $\Pi_\tau(\mathbf{a}) = (a_{i_1}, ..., a_{i_{n'}})$. If $A \subseteq \mathbb{F}_q^n$ is a set then its projection is denoted as a set $\widehat{\Pi}_\tau(A) = \{\Pi_\tau(\mathbf{a}) : \mathbf{a} \in A\}$. Let function $\beta : \mathbb{F}_q^l \to \mathbb{F}_2^l$ be such that for $\mathbf{a}$ from $\mathbb{F}_q^l$:

$$\beta(\mathbf{a}) = \mathbf{b}(\in \mathbb{F}_2^l) \text{ and } \operatorname{supp}(\mathbf{a}) = \operatorname{supp}(\mathbf{b}). \tag{3}$$

In order to generate matrix

$$\widetilde{G} = (\mathbf{e}_i)_{i=1}^l \tag{4}$$

of $[n, l, \widetilde{d}]$-code $\widetilde{C}$ and for vector $\mathbf{k}(\in \mathbb{F}_2^l)$ let us denote the submatrix of matrix (4) as $G_\mathbf{k}$ so that $G_\mathbf{k} = (\mathbf{e}_i)_{i \in \operatorname{supp}(\mathbf{k})}$. Random encoding parametrized with binary vector $\mathbf{k} \in \mathbb{F}_2^l$ is a function $g_\mathbf{k} : \mathbb{F}_q^{w(\mathbf{k})} \times \mathbb{F}_q^{l-w(\mathbf{k})} \to \widetilde{C}$ that

$$g_\mathbf{k}(\mathbf{m}, \mathbf{r}) = \mathbf{m}G_\mathbf{k} + \mathbf{r}G_{\overline{\mathbf{k}}} = \mathbf{c} \tag{5}$$

where $\mathbf{m} \in \mathbb{F}_q^{w(\mathbf{k})}$, $\mathbf{r}$ is chosen from $\mathbb{F}_q^{l-w(\mathbf{k})}$ randomly and equiprobably, $\overline{\mathbf{k}} = \mathbf{1} \oplus \mathbf{k}$, $\mathbf{1} \in \mathbb{F}_2^l$ and $w(\mathbf{1}) = l$. Let $\mathbf{c}'$ be a partially erased vector corresponding to the code block $\mathbf{c}$ (see (5)). If $\mathbf{k}$ *is known* then one can try to extract informational vector $\mathbf{m}'$ with the next rule:

$$\mathbf{m}' = g_\mathbf{k}^{-1}(\mathbf{c}') = \Pi_{\operatorname{supp}(\mathbf{k})}\left(\operatorname{Dec}_{\widetilde{C}}(\mathbf{c}')\right). \tag{6}$$

The set of all possible functions $g_\mathbf{k}$ for given $\widetilde{G}$ we denote as $\mathscr{G}(\widetilde{G})$:

$$\mathscr{G}(\widetilde{G}) = \{g_\mathbf{k} : \mathbf{k} \in \mathbb{F}_2^l\}.$$

The legal sender (or the user who has a permission to write symbols of code blocks into distributed storage) chooses function $g_\mathbf{k}$ randomly and equiprobably from $\mathscr{G}(\widetilde{G})$. With this assumption the legal receiver (the user who has a permission to read data from distributed storage) is not able to recover $\mathbf{m}'$ uniquely with only one code block $\mathbf{c}'$ because he or she has to know the set of coordinates in $\operatorname{Dec}_{\widetilde{C}}(\mathbf{c}')$ corresponding to the informational vector (see (6)). The legal receiver should know $\mathbf{k}$ for recovering the informational block. We propose to put the information about vector $\mathbf{k}$ into a package of $\theta + 1$ code blocks, $\theta \in \mathbb{N}$. Note that it is usual for data storage systems to read and write data as packages of blocks rather than single blocks.

Let us consider how the legal sender forms t-th package, $t \in \mathbb{N}$. The data from the source are represented as packages of θ blocks. The length of blocks may be different in different packages. At first, the sender gets vector $\mathbf{k}(\in \mathbb{F}_2^l)$ randomly and equiprobably. This vector is matched with function $g_\mathbf{k}$. At the next step the sender represents the data as a sequence of θ blocks with the length equal to $w(\mathbf{k})$ so that t-th package $\mathbf{M}_t$ of informational vectors is

$$\mathbf{M}_t = (\mathbf{m}_{t,1}, ..., \mathbf{m}_{t,\theta}), \quad \mathbf{m}_{t,1} \in \mathbb{F}_q^{w(\mathbf{k})}.$$

The corresponding package of code blocks is

$$\mathbf{C}_t = (\mathbf{c}_{t,1}, ..., \mathbf{c}_{t,\theta}, \mathbf{c}_{t,\theta+1}) \tag{7}$$

where $\mathbf{c}_{t,p} = g_\mathbf{k}(\mathbf{m}_{t,p}, \mathbf{r}_{t,p})$ for $p = 1, ..., \theta$ and $\mathbf{c}_{t,\theta+1} = g_\mathbf{k}(\mathbf{u}_t, \mathbf{0})$, $\mathbf{0}$ is a zero vector, $w(\mathbf{u}_t) = w(\mathbf{k})$. Vector $\mathbf{u}_t$ is chosen from the set of vectors with weight $w(\mathbf{k})$ $(\mathbf{u}_t \in \mathbb{F}_q^{w(\mathbf{k})})$ with probability equal to $(q-1)^{-w(\mathbf{k})}$ for every vector. Let us denote encoding of $\mathbf{M}_t$ as

$\mathrm{Enc}(\mathbf{M}_t) = \mathbf{C}_t$. The legal receiver can use the following way for extracting informational blocks from the packages. He or she should calculate vector $\mathbf{k}' = \beta(\mathrm{Dec}_{\widetilde{C}}(\mathbf{c}'_{t,\theta+1}))$ and then find $\mathbf{m}'_{t,p} = g^{-1}_{\mathbf{k}'}(\mathbf{c}'_{t,p})$, $p = 1, ..., \theta$. We denote decoding of package $\mathbf{C}'_t$ as $\mathrm{Dec}(\mathbf{C}'_t) = \mathbf{M}'_t = \{\mathbf{m}'_{t,1}, ..., \mathbf{m}'_{t,\theta}\}$. We denote constructed modification of the code noising method as $(\mathscr{G}(\widetilde{G}), \theta)$-scheme.

For our method the code rate is equal to $R_{\mathbf{k},\theta} = \frac{\theta w(\mathbf{k})}{(\theta+1)n} R$ for fixed $\mathbf{k}(\in \mathbb{F}_2^l)$ where $R = l/n$ is the code rate for code $\widetilde{C}$. As vector $\mathbf{k}$ is chosen randomly and equiprobably, the expected value R_θ of code rate is

$$R_\theta = \sum_{\mathbf{k}\in\mathbb{F}_2^l} \frac{\theta w(\mathbf{k})}{(\theta+1)n2^l} R = \frac{\theta}{(\theta+1)n2^l} \sum_{\mathbf{k}\in\mathbb{F}_2^l} w(\mathbf{k}) R = \frac{\theta}{2(\theta+1)} R \tag{8}$$

as $\sum_{\mathbf{k}\in\mathbb{F}_2^l} w(\mathbf{k}) = n2^{l-1}$. Note that $\lim_{\theta\to\infty} R_\theta = 0,5R$.

Let us denote the set $\mathscr{G}(\widetilde{G})_{h_1,h_2} = \{g_{\mathbf{k}} : h_1 \leqslant w(\mathbf{k}) \leqslant h_2\}$ for $h_1, h_2 \in \{0, ..., l\}$, $h_2 \geqslant h_1$. One may use this set if it is necessary to increase the code rate R_θ, for example. Note that $\mathscr{G}(\widetilde{G}) = \mathscr{G}(\widetilde{G})_{0,l}$. In the next section the $(\mathscr{G}(\widetilde{G}), \theta)$-scheme is analyzed for resistance against failure of $n - \nu$ repositories and coalition of μ participants (recall that here the length n of code block is equal to the number of repositories). It is easy to generalize the results represented in the next section if $(\mathscr{G}(\widetilde{G})_{h_1,h_2}, \theta)$-scheme is used instead of $\mathscr{G}(\widetilde{G})$.

2. ANALYSIS AND APPLICATION OF $(\mathscr{G}(\widetilde{G}), \theta)$-SCHEME

2.1. Defense against unreliable nodes

Theorem 1. *Let $\widetilde{C}$ be a $[n, l, \widetilde{d}]$-code generating matrix $\widetilde{G}$, and package $\mathbf{C}_t = \mathrm{Enc}(\mathbf{M}_t)$ be an output of $(\mathscr{G}(\widetilde{G}), \theta)$-scheme using function $g_{\mathbf{k}}$, $\mathbf{C}'_t = (\mathbf{c}'_{t,1}, ..., \mathbf{c}'_{t,\theta+1})$ is the corresponding package of partially erased code blocks. If every block $\mathbf{c}'_{t,p}$, $p = 1, ..., \theta + 1$ has no more than $\widetilde{d} - 1$ erasures, then $\mathrm{Dec}(\mathbf{C}'_t) = \mathbf{M}_t$.*

Proof. By condition, $\widetilde{d}$ is code distance of code $\mathscr{C}$ and there are no more than $\widetilde{d} - 1$ erasures in every code block. Then $\beta(\mathrm{Dec}_{\widetilde{C}}(\mathbf{c}'_{t,\theta+1})) = \mathbf{k}$. According to condition of the theorem, $g^{-1}_{\mathbf{k}}(\mathbf{c}'_{t,p}) = \mathbf{m}_{t,p}$ for $p = 1, ..., \theta$. □

Theorem 1 allows us to get limit on number $n - \nu$ of unreliable nodes when these nodes may be inaccessible but $(\mathscr{G}(\widetilde{G}), \theta)$-scheme provides the recovery of information. Exactly, $n - \nu \leqslant \widetilde{d} - 1$ where $\widetilde{d}$ is the code distance of $\widetilde{C}$.

2.2. Defense against coalition of untrusted nodes

If the coalition (or eavesdropper) knows function $g_{\mathbf{k}}$ then the resistance of the modified code noising method does not exceed the resistance of classical code noising based on the pair $(\widetilde{C}, \mathscr{L}(G_{\overline{\mathbf{k}}}))$ where $\mathscr{L}(A)$ is a linear subspace with rows of matrix A as its basis. In other words if adversary knows $\mathbf{k}$ he or she will be able to attack with all known ways, e.g. attack on repetitive messages. Further we presume that the next hypothesis is right.

Hypothesis 1. *If someone wants to get any information about data in package (7) he or she should obtain information about function $g_{\mathbf{k}}$ which was used while package encoding.*

Let K be a vector chosen randomly and equiprobably from $\mathbb{F}_2^l$, U be a random vector with distribution

$$p_U(\mathbf{u}) = \frac{1}{2^l(q-1)^{w(\mathbf{u})}} \tag{9}$$

on $\mathbb{F}_q^l$. Obviously, random vectors K and $\beta(U)$ have the same distributions. Let us consider for a fixed $\mathbf{k}$ the random vector

$$C^{\mathbf{k}} = g_{\mathbf{k}}(M^{(w(\mathbf{k}))}, R^{(l-w(\mathbf{k}))}),$$

where $M^{(w(\mathbf{k}))}$ and $R^{(l-w(\mathbf{k}))}$ are random vectors distributed uniformly over $\mathbb{F}_q^{w(\mathbf{k})}$ and $\mathbb{F}_q^{l-w(\mathbf{k})}$ respectively. Note that random vector $C^{\mathbf{k}}$ has uniform distribution over $\widetilde{C}$ for any $\mathbf{k}$. Let $\mathrm{H}(K)$ and $\mathrm{H}(K|C^{\mathbf{k}})$ be an entropy of a random vector K and conditional entropy of a random vector K on condition $C^{\mathbf{k}}$ respectively:

$$\mathrm{H}(K) = -\sum_{\chi \in \mathbb{F}_2^l} p_K(\chi)\log_2(p_K(\chi)) = l,$$

$$\mathrm{H}(K|C^{\mathbf{k}}) = -\sum_{\chi \in \mathbb{F}_2^l}\sum_{\mathbf{c}\in\widetilde{\mathbf{C}}} p_{(K,C^{\mathbf{k}})}(\chi,\mathbf{c})\log_2(p_{K|C^{\mathbf{k}}}(\chi|\mathbf{c})).$$

If $C^{\mathbf{k}} = \mathbf{c}(\in \widetilde{C})$, there is no way to choose a correct function from $\mathscr{G}(\widetilde{G})$ using only decoded value $\mathrm{Dec}_{\widetilde{C}}(\mathbf{c})$ because vectors $M^{(w(\mathbf{k}))}$, $R^{(l-w(\mathbf{k}))}$, and K are distributed uniformly. Thus $\mathrm{H}(K|C^{\mathbf{k}}) = l$ and the mutual information $\mathrm{I}(K;C^{\mathbf{k}})$ is equal to zero:

$$\mathrm{I}(K;C^{\mathbf{k}}) = \mathrm{H}(K) - \mathrm{H}(K|C^{\mathbf{k}}) = 0.$$

Moreover, it is not hard to check that $\mathrm{I}(K;(C_1^{\mathbf{k}},...,C_\theta^{\mathbf{k}})) = 0$ for θ copies $C_1^{\mathbf{k}},...,C_\theta^{\mathbf{k}}$ of a random vector $C^{\mathbf{k}}$.

Let us consider the random vectors C_1,...,C_θ, $X = U\widetilde{G}$, $C_p = C_p^{\beta(\mathbf{u})}$ if $U = \mathbf{u}$, $p = 1,...,\theta$, where U is a random vector with distribution (9). Let $\tau(\subseteq \underline{n})$ be a set of observed coordinates (or the numbers of repositories from the coalition) with cardinality $|\tau| = \mu$ and Z_1,...,Z_θ, Y be random vectors, $Z_p = \Pi_\tau(C_p)$, $p = 1,...,\theta$, $Y = \Pi_\tau(X)$. It is not hard to check the next chain of equalities:

$$\Pr\{\beta(U)=\mathbf{k}|Z_1=\mathbf{z}_1,...,Z_\theta=\mathbf{z}_\theta,Y=\mathbf{y}\}=\frac{\Pr\{\beta(U)=\mathbf{k},Z_1=\mathbf{z}_1,...,Z_\theta=\mathbf{z}_\theta,Y=\mathbf{y}\}}{\Pr\{Z_1=\mathbf{z}_1,...,Z_\theta=\mathbf{z}_\theta,Y=\mathbf{y}\}}=$$

$$=\frac{\Pr\{\beta(U)=\mathbf{k},Z_1=\mathbf{z}_1,...,Z_\theta=\mathbf{z}_\theta\}\Pr\{Y=\mathbf{y}|\beta(U)=\mathbf{k},Z_1=\mathbf{z}_1,...,Z_\theta=\mathbf{z}_\theta\}}{\Pr\{Z_1=\mathbf{z}_1,...,Z_\theta=\mathbf{z}_\theta,Y=\mathbf{y}\}}=$$

$$=\frac{\Pr\{\beta(U)=\mathbf{k}\}\Pr\{Z_1=\mathbf{z}_1,...,Z_\theta=\mathbf{z}_\theta|\beta(U)=\mathbf{k}\}\Pr\{Y=\mathbf{y}|\beta(U)=\mathbf{k},Z_1=\mathbf{z}_1,...,Z_\theta=\mathbf{z}_\theta\}}{\Pr\{Z_1=\mathbf{z}_1,...,Z_\theta=\mathbf{z}_\theta\}\Pr\{Y=\mathbf{y}|Z_1=\mathbf{z}_1,...,Z_\theta=\mathbf{z}_\theta\}}=$$

$$=\frac{\Pr\{\beta(U)=\mathbf{k}\}\Pr\{Y=\mathbf{y}|\beta(U)=\mathbf{k}\}}{\Pr\{Y=\mathbf{y}\}}=\Pr\{\beta(U)=\mathbf{k}|Y=\mathbf{y}\}=\Pr\{K=\mathbf{k}|Y=\mathbf{y}\},$$

for every $\mathbf{k}\in\mathbb{F}_2^k$, $\mathbf{z}_1$,..., $\mathbf{z}_\theta$, $\mathbf{y}\in\Pi_\tau(\widetilde{C})$. Thus, $\mathrm{H}(K|Z_1,...,Z_\theta,Y) = \mathrm{H}(K|Y)$ and

$$\mathrm{I}(K;(Z_1,...,Z_\theta,Y)) = H(K) - \mathrm{H}(K|Y) = \mathrm{I}(K;Y) = l - \mathrm{H}(K|Y), \tag{10}$$

because K is equiprobable. For every $\mathbf{k}\in\mathbb{F}_2^l$ we denote $B(\mathbf{k}) = \{\mathbf{u}\in\mathbb{F}_q^l : \beta(\mathbf{u}) = \mathbf{k}\}$. Let $n\in\mathbb{N}$, $\tau\subseteq\underline{n}$, $\mathbf{y}$ be the implementation of a random vector $\mathbf{Y} = \Pi_\tau(\mathbf{X})$. Consider the the system of equations

$$\mathbf{u}\Pi_\tau(\widetilde{G}) = \mathbf{y} \tag{11}$$

where $\mathbf{u}$ is unknown. The set of solutions of this system denote $\Gamma(\mathbf{y})$.

Lemma 1. *Let* $\hat{\beta}(\Gamma(\mathbf{y})) = \{\beta(\mathbf{g}) : \mathbf{g} \in \Gamma(\mathbf{y})\}$. *Then*

$$\mathrm{H}(K|Y=\mathbf{y}) \leqslant \log_2 |\hat{\beta}(\Gamma(\mathbf{y}))| \leqslant \min\{(l - \mathrm{rank}(\Pi_\tau(\widetilde{G}))) \log_2 q; l\}.$$

Proof. Note that $p_{U|Y}(\mathbf{u}|\mathbf{y}) = 0$ if $\mathbf{u} \notin \Gamma(\mathbf{y})$. Then

$$p_{K|Y}(\mathbf{k}|\mathbf{y}) = \sum_{\mathbf{u}\in B(\mathbf{k})} p_{U|Y}(\mathbf{u}|\mathbf{y}) =$$

$$= \sum_{\mathbf{u}\in B(\mathbf{k})\cap\Gamma(\mathbf{y})} \frac{p_U(\mathbf{u})}{\sum_{\mathbf{u}'\in\Gamma(\mathbf{y})} p_U(\mathbf{u}')} = \frac{\sum_{\mathbf{u}\in B(\mathbf{k})\cap\Gamma(\mathbf{y})} 2^{-l}(q-1)^{-w(\mathbf{u})}}{\sum_{\mathbf{u}'\in\Gamma(\mathbf{y})} 2^{-l}(q-1)^{-w(\mathbf{u}')}} =$$

$$= \frac{\sum_{\mathbf{u}\in B(\mathbf{k})\cap\Gamma(\mathbf{y})}(q-1)^{-w(\mathbf{k})}}{\sum_{\mathbf{u}\in\Gamma(\mathbf{y})}(q-1)^{-w(\mathbf{u})}} = \frac{|B(\mathbf{k})\cap\Gamma(\mathbf{y})|}{\sum_{\mathbf{u}\in\Gamma(\mathbf{y})}(q-1)^{w(\mathbf{k})-w(\mathbf{u})}}. \qquad (12)$$

It is obvious that

$$p_{K|Y}(\mathbf{k}|\mathbf{y}) \neq 0 \Leftrightarrow B(\mathbf{k})\cap\Gamma(\mathbf{y}) \neq \varnothing \Leftrightarrow \mathbf{k} \in \hat{\beta}(\Gamma(\mathbf{y})),$$

then

$$\mathrm{H}(K|Y=\mathbf{y}) = -\sum_{\mathbf{k}\in\hat{\beta}(\Gamma(\mathbf{y}))} p_{K|Y}(\mathbf{k}|\mathbf{y}) \log_2(p_{K|Y}(\mathbf{k}|\mathbf{y})) =$$

$$= -\sum_{\mathbf{k}\in\hat{\beta}(\Gamma(\mathbf{y}))} \frac{|B(\mathbf{k})\cap\Gamma(\mathbf{y})|}{\sum_{\mathbf{u}\in\Gamma(\mathbf{y})}(q-1)^{w(\mathbf{k})-w(\mathbf{u})}} \log_2\left(\frac{|B(\mathbf{k})\cap\Gamma(\mathbf{y})|}{\sum_{\mathbf{u}\in\Gamma(\mathbf{y})}(q-1)^{w(\mathbf{k})-w(\mathbf{u})}}\right) \leqslant \log_2 |\hat{\beta}(\Gamma(\mathbf{y}))|,$$

because $\log_2 |\hat{\beta}(\Gamma(\mathbf{y}))|$ is the entropy of uniformly distributed K for a given $\mathbf{y}$. Estimate of $\log_2 |\hat{\beta}(\Gamma(\mathbf{y}))|$ is also right because there are only 2^l possible variants of vector $\mathbf{k}(\in \mathbb{F}_2^l)$, on the one hand, and equation (11) has $q^{l-\mathrm{rank}(\Pi_\tau(\widetilde{G}))}$ solutions, on the other hand. □

Let $B(\mathbf{k},\mathbf{y}) = B(\mathbf{k})\cap\Gamma(\mathbf{y})$ and for $i \in \{0,...,l\}$ define $A_i(\Gamma(\mathbf{y})) = \sum_{\mathbf{u}\in\Gamma(\mathbf{y})}(q-1)^{i-w(\mathbf{u})}$. Then from (12) we get $p_{K|Y}(\mathbf{k}|\mathbf{y}) = \frac{|B(\mathbf{k},\mathbf{y})|}{A_{w(\mathbf{k})}(\Gamma(\mathbf{y}))}$,

$$\mathrm{H}(K|Y=\mathbf{y}) = -\sum_{\mathbf{k}\in\mathbb{F}_2^l} p_{K|Y}(\mathbf{k}|\mathbf{y}) \log_2 p_{K|Y}(\mathbf{k}|\mathbf{y}) = -\sum_{\mathbf{k}\in\mathbb{F}_2^l} \frac{|B(\mathbf{k},\mathbf{y})|}{A_{w(\mathbf{k})}(\Gamma(\mathbf{y}))} \log_2\left(\frac{|B(\mathbf{k},\mathbf{y})|}{A_{w(\mathbf{k})}(\Gamma(\mathbf{y}))}\right) =$$

$$= -\sum_{i=0}^{l} \frac{1}{A_i(\Gamma(\mathbf{y}))} \sum_{\mathbf{k}\in\mathbb{F}_2^l : w(\mathbf{k})=i} |B(\mathbf{k},\mathbf{y})| \log_2\left(\frac{|B(\mathbf{k},\mathbf{y})|}{A_i(\Gamma(\mathbf{y}))}\right) =$$

$$= -\sum_{i=0}^{l} \frac{1}{A_i(\Gamma(\mathbf{y}))} \left[\sum_{\mathbf{k}\in\mathbb{F}_2^l : w(\mathbf{k})=i} |B(\mathbf{k},\mathbf{y})| \left(\log_2(|B(\mathbf{k},\mathbf{y})|) - \log_2(A_i(\Gamma(\mathbf{y})))\right)\right] =$$

$$= -\sum_{i=0}^{l} \frac{1}{A_i(\Gamma(\mathbf{y}))} \left[\sum_{\mathbf{k}\in\mathbb{F}_2^l : w(\mathbf{k})=i} |B(\mathbf{k},\mathbf{y})| \log_2(|B(\mathbf{k},\mathbf{y})|) - \log_2(A_i(\Gamma(\mathbf{y}))) \sum_{\mathbf{k}\in\mathbb{F}_2^l : w(\mathbf{k})=i} |B(\mathbf{k},\mathbf{y})|\right] =$$

$$= -\sum_{i=0}^{l} \frac{1}{A_i(\Gamma(\mathbf{y}))} \left[\sum_{\mathbf{k}\in\mathbb{F}_2^l : w(\mathbf{k})=i} |B(\mathbf{k},\mathbf{y})| \log_2(|B(\mathbf{k},\mathbf{y})|) - N_i(\Gamma(\mathbf{y})) \log_2(A_i(\Gamma(\mathbf{y})))\right],$$

where $N_i(\Gamma(\mathbf{y})) = |\{\mathbf{u} \in \Gamma(\mathbf{y}) : w(\mathbf{u}) = i\}|$. Thus, if $q \neq 2$, then calculation of $\mathrm{H}(K|Y = \mathbf{y})$ and $\mathrm{I}(K;Y)$ seems to be a hard challenge. Because two or more different solutions of the system (11) can correspond to one binary vector $\mathbf{k}$ if their supports are the same. Next theorem calculates $\mathrm{I}(K;Y)$ for $q = 2$.

Theorem 2. *Let $q = 2$, $\tau \subseteq \underline{n}$, then $\mathrm{I}(K;Y) = \mathrm{rank}(\Pi_\tau(\widetilde{G}))$ for $(\mathscr{G}(\widetilde{G}), \theta)$-scheme.*

Proof. As $q = 2$ then $|B(\mathbf{k})| = 1$ and $\Gamma(\mathbf{y}) = \hat{\beta}(\Gamma(\mathbf{y}))$, because different binary vectors have different supports. If $\mathbf{y}$ is fixed, then using Lemma 1 for $\mathbf{k} \in \Gamma(\mathbf{y})$ we have:

$$p_{K|Y}(\mathbf{k}|\mathbf{y}) = \frac{|B(\mathbf{k}) \cap \Gamma(\mathbf{y})|}{\sum_{\mathbf{u} \in \Gamma(\mathbf{y})} (2-1)^{w(\mathbf{k})-w(\mathbf{u})}} = \frac{1}{|\Gamma(\mathbf{y})|},$$

$$\mathrm{H}(K|Y = \mathbf{y}) = -\sum_{\mathbf{k} \in \Gamma(\mathbf{y})} \frac{1}{|\Gamma(\mathbf{y})|} \log_2 (\frac{1}{|\Gamma(\mathbf{y})|}) = \log_2 |\Gamma(\mathbf{y})|.$$

As $|\Gamma(\mathbf{y})| = 2^{l - \mathrm{rank}(\Pi_\tau(\widetilde{G}))}$, then $\mathrm{H}(K|Y) = \mathrm{H}(K|Y = \mathbf{y}) = l - \mathrm{rank}(\Pi_\tau(\widetilde{G}))$. The last step is the substitution of the value of $\mathrm{H}(K|Y)$ into (10). □

It follows from (10) and Lemma 1 that obtaining information $\mathrm{I}(K; (Z_1, ..., Z_\theta, Y = \mathbf{y})) = \mathrm{I}(K; Y = \mathbf{y})$ is strongly related with constructing $\Gamma(\mathbf{y})$ as the set of solutions of the system (11). As $|\Gamma(\mathbf{y})| = q^{l - \mathrm{rank}(\Pi_\tau(\widetilde{G}))}$, this task can be challenging in selecting parameters of the scheme. In general, if $(\mathscr{G}(\widetilde{G})_{h_1,h_2}, \theta)$ is used, then

$$\mathrm{I}(K; Y = \mathbf{y}) \geqslant \log_2(|\mathscr{G}(\widetilde{G})_{h_1,h_2}|) - \log_2(\hat{\beta}(\Gamma(\mathbf{y})) \cap \{\mathbf{k} \in \mathbb{F}_2^l : h_1 \leqslant w(\mathbf{k}) \leqslant h_2\}).$$

The complexity of obtaining this information seems to be equivalent to the complexity of making the set

$$\hat{\beta}(\Gamma(\mathbf{y})) \cap \{\mathbf{k} \in \mathbb{F}_2^l : h_1 \leqslant w(\mathbf{k}) \leqslant h_2\}.$$

To make this set it is necessary either to construct the set $\Gamma(\mathbf{y})$ or to look over all possible vectors from $\mathbb{F}_q^l$, then choose vectors with weight in range $[h_1, h_2]$ only and check if these vectors are solutions of the system (11). Furthermore, if $|\mathscr{G}(\widetilde{G})_{h_1,h_2}|$ is small, the eavesdropper is able to check *all* functions from $\mathscr{G}(\widetilde{G})_{h_1,h_2}$. Thus, the computational complexity of obtaining information *about package of informational blocks* (when Hypothesis 1 is right) is not less than

$$\mathscr{O}\left(\min\left\{|\mathscr{G}(\widetilde{G})_{h_1,h_2}|, |\Gamma(\mathbf{y})|\right\}\right), \tag{13}$$

where $\mathscr{O}(|\mathscr{G}(\widetilde{G})_{h_1,h_2}|)$ is the complexity of brute force over all functions from $\mathscr{G}(\widetilde{G})_{h_1,h_2}$ and $\mathscr{O}(|\Gamma(\mathbf{y})|)$ is the complexity of making the set $\Gamma(\mathbf{y})$.

Note that for $q = 2$ obtaining full information even about the *length* of informational blocks by package (7) is a severe challenge. Consider the general case when $\mathscr{G}(\widetilde{G})_{h_1,h_2}$-scheme is used. Actually the eavesdropper will get non-zero information about the length only if there is at least one number $l' \in \{h_1, ..., h_2\}$ such that there is not any vector with weight l' in $\Gamma(\mathbf{y})$. The complexity of obtaining information about the length on conditions $\mu < l$ and $\mathrm{rank}(\Pi_\tau(\widetilde{G})) = \mu$ may be reduced to one task in the coding theory. Namely, the matrix $\Pi_\tau(\widetilde{G})$ may be considered as a transposed parity-check matrix of

some $[l, l-\mu]$-code. Let vector $\mathbf{y}$, number $w \in \{h_1, ..., h_2\}$, and transposed parity-check matrix $\Pi_\tau(\widetilde{G})$ be preassigned. The task of finding vector $\mathbf{u}$ with weight *no more than* w on condition (11) is NP-complete [12]. If $(\mathscr{G}(\widetilde{G})_{h_1,h_2}, \theta)$-scheme is used, then obtaining non-zero information about the length of message is equivalent to finding out that there is no vector $\mathbf{u}$ with weight *exactly* w on condition (11). We do not know any polynomial algorithm for resolving the latter task. Note that in binary case this problem is also NP-complete [13].

Thus, according to Hypothesis 1 the resistance of the modified code noising method to known attacks, particularly to the repetitive interception attack, is based on the fact that it may be difficult (depending on the parameters) for the coalition to obtain information about the mapping used. It should be noted that to increase resistance to repetitive interception attack it is also recommended to use small value of the parameter θ. In this case the probability of the appearance of code blocks, corresponding to one informational block with the same mapping, is reducing. If $\theta = 1$, then the level of defense is maximal in this sense, but code rate $R_\theta = 0,25$ is minimal.

We assume above that number of repositories is equal to the length of code block. It is practically unreal if length of code block is huge. But proposed $(\mathscr{G}(\widetilde{G}), \theta)$-scheme may be easily adopted for a smaller number of repositories. If N is the length of code block and n is the number of repositories ($n < N$), then we should write no more $\lceil N/n \rceil$ code symbols into every repository. In this case coalition with μ repositories knows no more than $\mu\lceil N/n \rceil$ symbols of every code block, and inaccessibility of $n - \nu$ repositories is equivalent to erasure of $(n-\nu)\lceil N/n \rceil$ symbols of every code block.

2.3. An example of $(\mathscr{G}(\widetilde{G}), \theta)$-scheme application

Let $\widetilde{C}$ be a $[255, 200, 56]$ Reed – Solomon code over $\mathbb{F}_{2^8}$, $q = 2^8$. The table contains the comparison of characteristics of $(\mathscr{G}(\widetilde{G}), \theta)$-scheme and the classical code noising method based on pair $(\widetilde{C}, C)$ if C is $[255, 150, 106]$ Reed – Solomon code. Code rate of a pair-based code noising method is $50/255 \approx 0.196$ and theoretical-informational resistance is achieved if coalition knows no more than 150 symbols of code block [1]. For $(\mathscr{G}(\widetilde{G}), \theta)$-scheme code rate is equal to 0.196 if $\theta = 1$ and is about 0.392 if $\theta \geqslant 1000$.

Comparison of characteristics of $(\mathscr{G}(\widetilde{G}), \theta)$-scheme and $(\widetilde{C}, C)$-pair

Number of repositories, n	3	5	17
length of a part, $\lceil N/n \rceil$	85	51	15
max. value $n - \nu$	0	1	3
max. value μ for $(\mathscr{G}(\widetilde{G}), \theta)$	**2**	**3**	**12**
max. value μ for $(\widetilde{C}, C)$	**1**	**2**	**10**

Maximal allowable size μ of coalition for $(\mathscr{G}(\widetilde{G}), \theta)$-scheme is calculated on condition that complexity (13) should be not less than 2^{128}; it corresponds to high level of resistance according to [14]. Note that estimation (13) for this example takes the form $\mathscr{O}(\min\{2^{200}, |\Gamma(\mathbf{y})|\})$. For any different sets τ_1 and τ_2 of the same cardinality for generating matrix of Reed – Solomon code we have $\operatorname{rank}(\Pi_{\tau_1}(\widetilde{G})) = \operatorname{rank}(\Pi_{\tau_2}(\widetilde{G}))$. Then for security reason the allowed maximal number x of symbols observed by the coalition in each code block can be obtained from inequality $2^{8(200-x)} \geqslant 2^{128}$; so, we have $x \leqslant 184$. Note that each repository knows only $\lceil \frac{255}{n} \rceil$ symbols of each code block where n is a number of repositories, $n \in \{3, 5, 17\}$. As we can see from the table, the modified method provides defense against a bigger coalition. In particular, in case of using three repositories the modified code noising method provides

computational resistance even in the case when two of the three participants have united in the coalition. At the same time the classical code noising method provides resistance only in the case of the coalition consisting of one participant.

CONCLUSION

Usually the core of the resistance of modern methods of data confidentiality protection is a certain mathematical problem with a computationally complex solution if particular "secret" is unknown. In this paper, a non-cryptographic method for protecting data confidentiality is constructed based on the use of special data coding and distribution of parts of the encoded data among the nodes of the distributed storage. In this case the "secret" is replaced with the assumption that the observer (i.e. node coalition) cannot get data from all nodes of the distributed storage. The paper shows that the complexity of recovering the protected data by coalition is not less than the complexity of solving the theoretical coding problem of finding all weights of vectors with a given syndrome. The computations lead to the conclusion that the constructed method can provide protection from coalitions of more cardinality than the classical code noising method, and provides not less protection from the failure of storage nodes.

References

1. Subramanian A., McLaughlin S. W. MDS codes on the erasure-erasure wiretap channel. *arXiv:0902.3286 [cs.IT]*, 2009.
2. Korzhik V., Yakovlev V. Nonasymptotic estimates of information protection efficiency for the wire-tap channel concept. In: *Seberry J., Zheng Y. (eds.). Advances in Cryptology — AUSCRYPT '92. AUSCRYPT 1992. Lecture Notes in Computer Science*, 1993, vol. 718, pp. 183–195. DOI: https://doi.org/10.1007/3-540-57220-1_61
3. Ozarov L. H., Wyner A. D. Wire-Tap Channel II. In: *Beth T., Cot N., Ingemarsson I. (eds.). Advances in Cryptology. EUROCRYPT 1984. Lecture Notes in Computer Science*, 1984, vol. 209, pp. 33–55. DOI: https://doi.org/10.1007/3-540-39757-4_5
4. Wei V. K. Generalized Hamming Weights for Linear Codes. *IEEE Trans. Inform. Theory*, 1991, vol. 37, no. 5, pp. 1412–1418. DOI: https://doi.org/10.1109/18.135655
5. Forney G. D. Dimension/Length Profiles and Trellis Complexity of Linear Block Codes. *IEEE Trans. Inform. Theory*, 1994, vol. 40, no. 6, pp. 1741–1752. DOI: https://doi.org/10.1109/18.340452
6. Luo Y., Mitrpant C., Hav Vinck A. J., Chen K. Some New characters on the wire-tap channel of type II. *IEEE Trans. Inform. Theory*, 2005, vol. 51, no. 3, pp. 1222–1229. DOI: https://doi.org/10.1109/TIT.2004.842763
7. Hu P., Sung C. W., Ho S.-W., Chan T. H. Optimal Coding and Allocation for Perfect Secrecy in Multiple Clouds. *IEEE Transactions on Information Forensics and Security*, 2016, vol. 11, no. 2, pp. 388–399. DOI: https://doi.org/10.1109/TIFS.2015.2500193
8. Kosolapov Yu. V. Codes for a generalized wire-tap channel model. *Problems of Information Transmission*, 2015, vol. 51, no. 1, pp. 20–24. DOI: https://doi.org/10.1134/S0032946015010020
9. Kosolapov Yu. V., Pozdnyakov A. V. Evaluation of resistance of code noising in the distributed data storage. *Systems and Means of Informatics*, 2015, vol. 25, no. 4, pp. 158–174 (in Russian). DOI: https://doi.org/10.14357/08696527150412
10. Gazaryan Yu. O., Kosolapov Yu. V. On the experimental estimation of the lower bound for the maximum number of messages in a scheme aimed at data protection against spoofing. *Computational Technologies*, 2015, vol. 20, no. 6, pp. 5–21 (in Russian).
11. Bellare M., Tessaro S., Vardy A. A Cryptographic Treatment of the Wiretap Channel. *arXiv:1201.2205 [cs.IT]*, 2012.

12. Barg S. Some new NP-complete coding problems. *Problems of Information Transmission*, 1994, vol. 30, no. 3, pp. 209–214.
13. Sendrier N., Simos D. E. The Hardness of Code Equivalence over $\mathbb{F}_q$ and Its Application to Code-Based Cryptography. In: *Gaborit P. (eds.). Post-Quantum Cryptography. PQCrypto 2013. Lecture Notes in Computer Science*, 2013, vol. 7932, pp. 203–216.
14. Lenstra A. K., Verheul E. R. Selecting Cryptographic Key Sizes. *J. Cryptology*, 2001, vol. 14, pp. 255–293. DOI: https://doi.org/10.1007/s00145-001-0009-4

Cite this article as:
Kosolapov Yu. V., Pevnev F. S. A Method of Protected Distribution of Data Among Unreliable and Untrusted Nodes. *Izv. Saratov Univ. (N. S.), Ser. Math. Mech. Inform.*, 2019, vol. 19, iss. 3, pp. 326–337. DOI: https://doi.org/10.18500/1816-9791-2019-19-3-326-337

УДК 621.391.7

Метод защищенного распределения данных среди ненадежных и недоверенных узлов

Ю. В. Косолапов, Ф. С. Певнев

Косолапов Юрий Владимирович, кандидат технических наук, доцент, Институт математики, механики и компьютерных наук имени И. И. Воровича, Южный федеральный университет, Россия, 344090, г. Ростов-на-Дону, ул. Мильчакова, д. 8а, itaim@mail.ru

Певнев Федор Сергеевич, магистрант, Институт математики, механики и компьютерных наук имени И. И. Воровича, Южный федеральный университет, Россия, 344090, г. Ростов-на-Дону, ул. Мильчакова, д. 8а, fes_21@mail.ru

В работе рассматривается модель защиты конфиденциальности и целостности данных в системе распределенного хранения. Предполагается, что информационные блоки кодируются в кодовые блоки, которые затем разделяются на части и распределяются среди узлов хранения распределенного хранилища. В качестве способа кодирования построена модификация метода кодового зашумления, которая одновременно обеспечивает вычислительную стойкость к коалиционным атакам на конфиденциальность хранимых данных, а также обеспечивает защиту от выхода из строя части узлов хранения. При этом защита конфиденциальности обеспечивается для коалиций бóльшей мощности, чем в случае применения классического метода кодового зашумления. Вычислительная стойкость основана на сложности решения одной теоретико-кодовой задачи.

Ключевые слова: канал с перехватом, защищенное распределенное хранилище, коалиционные атаки.

Поступила в редакцию: 05.10.2018 / Принята: 21.05.2019 / Опубликована: 31.08.2019

Образец для цитирования:
Kosolapov Yu. V., Pevnev F. S. A Method of Protected Distribution of Data Among Unreliable and Untrusted Nodes [Косолапов Ю. В., Певнев Ф. С. Метод защищенного распределения данных среди ненадежных и недоверенных узлов] // Изв. Сарат. ун-та. Нов. сер. Сер. Математика. Механика. Информатика. 2019. Т. 19, вып. 3. С. 326–337. DOI: https://doi.org/10.18500/1816-9791-2019-19-3-326-337