Journal of Mathematics and System Science 5 (2015) 516-522
doi: 10.17265/2159-5291/2015.12.004

Randomized Stream Ciphers with Enhanced Security Based on Nonlinear Random Coding

Anton Alekseychuk, Sergey Gryshakov
Institute of Special Communication and Information Security, National Technical University of Ukraine "KPI", Kiev, Ukraine

Received: September 07, 2015 / Accepted: October 05, 2015 / Published: December 25, 2015.

Abstract: We propose a framework for designing randomized stream ciphers with enhanced security. The key attribute of this framework is using of nonlinear bijective mappings or keyless hash functions for random coding. We investigate the computational security of the proposed ciphers against chosen-plaintext-chosen-initialization-vector attacks and show that it is based on the hardness of solving some systems of random nonlinear Boolean equations. We also provide guidelines for choosing components to design randomizers for specified ciphers.

Keywords: Symmetric cryptography, randomized stream cipher, random coding, computational security, chosen-plaintext-chosen-initialization-vector attack.

1. Introduction

In [1-5], a generic class of randomized stream ciphers based on joint employment of dedicated random (or homophonic) coding and error-correction coding by linear binary codes is proposed and studied. One of the goals of designing such ciphers is to increase the security (without substantial performance reducing) of stream ciphers currently used in wireless communication systems, particularly, in the GSM standard. Another reason is to construct symmetric encryption schemes, whose security can be reduced to the hardness of some known mathematical problem such as the problem of decoding a random linear code [6].

Further investigation of the ciphers proposed in [3, 4] showed [7] that their computational security significantly depends on the properties of their components and can be considerably less than their designers claim. In particular, some of specified ciphers are vulnerable even to ciphertext-only attacks and the problem of choosing components for their design (according to both security and practicality requirements) is non-trivial and requires further research.

In this paper, we propose another framework for designing randomized stream ciphers with enhanced security. This framework is based on ideas from [8, 9] and consists in using of nonlinear bijective mappings or keyless hash functions for random coding. It is shown that the security of the proposed ciphers in the chosen-plaintext-chosen-initialization-vector (IV) attacking scenario is based on the hardness of solving some systems of random nonlinear Boolean equations. In addition, we provide guidelines for choosing components to design randomizers for specified ciphers.

2. Proposed Framework

For any natural n denote by V_n the set of all n-dimensional Boolean vectors.

The initial objects for a randomized stream cipher with parameters $l, m \in \mathbf{N}$, where $l < m$, and a key space K are:

Corresponding author: Anton Alekseychuk, Doctor of Technical Science, Professor of Institute of Special Communication and Information Security of NTUU "KPI", research field: theoretical cryptography, E-mail: alex-dtn@ukr.net.

(1) a mapping $\phi: V_{m-l} \to V_l$;

(2) a commutative group operation $*$ on the set V_m;

(3) a permutation matrix P of order m;

(4) a keystream generator that produces a sequence $f_0(k), f_1(k), \ldots$ of m-dimensional Boolean vectors determined by a key $k \in K$. It is assumed that the functions $f_i: K \to V_m$, $i = 0, 1, \ldots$, can depend on some public parameters (initialization vectors).

To encrypt a plaintext $s_0, s_1, \ldots, s_t$, where $s_i \in V_l$, $i = 0, 1, \ldots, t$, with a key $k \in K$ the sender generates a sequence of independent random vectors $u_0, u_1, \ldots, u_t$, where u_i is uniformly distributed on the set V_{m-l}, and computes the ciphertext $z_0, z_1, \ldots, z_t$ as follows:

$$z_i = (u_i, s_i \oplus \phi(u_i))P * f_i(k), \quad i = \overline{0,t}, \qquad (1)$$

where $\oplus$ denotes the bitwise XOR operation. The legitimate receiver, knowing $f_i(k)$, can find the message $(z_{1,i}, z_{2,i}) = z_i *^{-1} f_i(k)$, where $z_{1,i} \in V_{m-l}$, $z_{2,i} \in V_l$, and the operation $*^{-1}$ is defined by the relation: $x = y *^{-1} z \Leftrightarrow y = x * z$, $x, y, z \in V_m$. After that he can recover s_i by the formula $s_i = \phi(z_{1,i}) \oplus z_{2,i}$ (see Fig. 1). On the other hand, the adversary in order to find the key k will be forced to deal with a corrupted keystream

$$(u_i, s_i \oplus \phi(u_i))P * f_i(k), \quad i = \overline{0,t}.$$

Let us remark that the objects ϕ, $*$, P should be chosen under the requirements for both the cryptographic security and the implementation efficiency of transformation (1). Taking into account the last requirement, we can set, for example, $a * b = (a + b) \bmod 2^m$, where arbitrary vectors $a, b \in V_m$ are identified with the corresponded numbers in the set $\{0, 1, \ldots, 2^m - 1\}$, and define P as the matrix of a rotation by a certain number of bits. The mapping ϕ should be chosen much more carefully because its properties influence essentially on the security of the considered cipher (see details below).

We propose to use one of the two general approaches:

(1) use as ϕ a bijective mapping on the set V_l (for $m = 2l$) "with good cryptographic properties" such as those used in modern block ciphers;

(2) use as ϕ a keyless hash function (such as Keccak [10]).

Taking into account the fact that a secure hash function simulates a random mapping (in our case from V_{m-l} to V_l) sufficiently well, the last variant looks more preferable with regard to providing adequate security of the randomized cipher.

Thus the key attribute of the proposed framework is using of the above mentioned nonlinear mappings for designing randomized stream ciphers. This is the main difference from the framework described in [1 – 5], where only binary linear transformations, particularly, error-correction coding of input messages by linear codes, are used. (Emphasize that the mentioned coding is not used in (1) at all).

Also note that, in comparison with randomized block ciphers [8, 9], nonlinear mappings used in randomizers of the proposed stream ciphers have slightly different requirements related to the specific attacks precisely on stream ciphers. This determines the differences between design criteria for randomizers used in randomized block and stream ciphers, respectively. In particular, keyless hash functions can be used in randomized stream ciphers with nonlinear random coding that differs the proposed framework from the one described in [8, 9].

3. Security Evaluation of the Proposed Ciphers Against Chosen-plaintext-chosen-IV Attacks

3.1 Basic Attack

Let's consider one of the most powerful attacks on randomized stream ciphers [7], when the adversary has access to the encryption oracle with unknown (chosen uniformly at random from the set K) key k

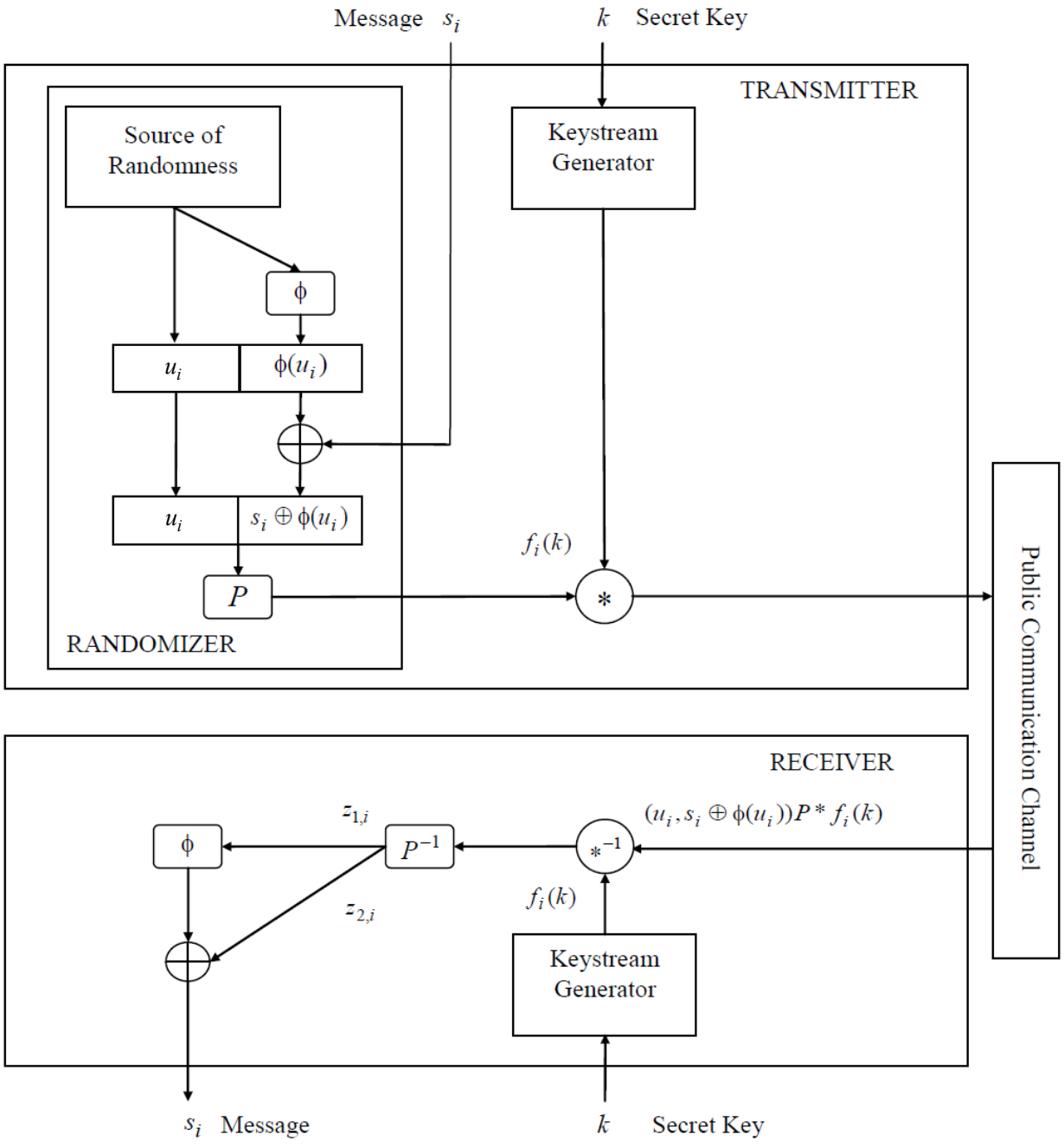

Fig. 1 Block diagram of proposed randomized stream cipher

and can choose on his own initialization vectors determining the functions f_i, $i = 0, 1, \ldots$. The aim of the attack is to recover for some fixed i the vector $f_i(k)$ from a collection of messages obtained by encrypting t times the same message $s_i = 0$ under the same IV. In this case, the adversary can derive equations of the form

$$(u_j, \phi(u_j))P * f_i(k) = y_j, \quad j = \overline{1, t},$$

where $y_1, y_2, \ldots, y_t$ are known and $f_i(k)$, $u_1, \ldots, u_t$ are not.

Now consider a more general problem.

Let

$$\xi_j + x = y_j, \quad j = \overline{1, t} \qquad (2)$$

be a system of random equations over a finite abelian group $(G,+)$, where $\xi_1, \xi_2, ..., \xi_t$ are independent random variables with uniform distribution on a set $M \subseteq G$, $y_j = x_0 + \xi_j$ is the result of substitution an unknown element $x_0 \in G$ into j-th equation of the system, $j = \overline{1,t}$. It is required to recover this element from the known values $y_1, y_2, ..., y_t$ and M.

It is obvious that recovery of the vector $f_i(k)$ in the considered attacking scenario is reduced to the solving the above formulated problem if

$$(G,+) = (V_m, *), \quad x_0 = f_i(k),$$

$$M = \{(u, \phi(u))P : u \in V_{m-l}\}. \qquad (3)$$

It is also clear that the set of all solutions of system (2) is equal to $\bigcap_{j=1}^{t}(y_j - M)$ and contains, at least, one element (equal to x_0).

For any $x \in G$ denote $S_x(\xi) = \bigcap_{j=1}^{t}(x + \xi_j - M)$, where $\xi = (\xi_1, \xi_2, ..., \xi_t)$. Then $S_x(\xi)$ is the intersection of independent random sets distributed as follows:

$$\mathbf{P}\{x + \xi_j - M = A\} = \frac{|\{y \in M : A = x + y - M\}|}{|M|},$$
$$A \subseteq G.$$

To find the solution x_0 of system (2) the adversary can use the following most natural algorithm.

Algorithm 1: *exhaustive search over the values of* ξ_1 *and checking the condition* $y_1 - \xi_1 \in \bigcap_{j=2}^{t}(y_j - M)$ *(it is assumed that searching is executed until the first success).*

Let's evaluate the time complexity of Algorithm 1. Suppose that the addition of any two elements $a, b \in G$ and the check of condition $a \in M$ for any $a \in G$ take constant time.

Let's denote

$$d_M(a) = \frac{|\{y \in M : y + a \in M\}|}{|M|}, \quad a \in G, \qquad (4)$$

$$d_M = \max\{d_M(a) : a \in G \setminus \{0\}\}. \qquad (5)$$

Statement 1. *Suppose that* $d_M < 1$. *Then for any* $\delta \in (0,1)$ *and*

$$t = \left\lceil \frac{\log(\delta^{-1}|M|)}{\log({d_M}^{-1})} \right\rceil + 1$$

the solution x_0 *of system (2) can be found with probability at least* $1 - \delta$ *in time* $O(|M|t)$.

Proof. It is sufficient to prove that the error probability, p_e, of Algorithm 1 satisfies the inequality

$$p_e \le |M| {d_M}^{t-1}. \qquad (6)$$

Suppose that Algorithm 1 makes a mistake; then there exists an element $x \in G \setminus \{x_0\}$, which belongs to the set $S_{x_0}(\xi) = \bigcap_{j=1}^{t}(x_0 + \xi_j - M)$. Since $\xi_1, \xi_2, ..., \xi_t$ are independent random variables with uniform distribution on the set M we get

$$p_e \le \sum_{x \ne x_0} \mathbf{P}\{x \in S_{x_0}(\xi)\} =$$

$$= \sum_{x \ne x_0} \left(|M|^{-1} \cdot |\{y \in M : y + x_0 - x \in M\}| \right)^t.$$

Now, using (4), (5) and the notation $I_M(z)$, $z \in G$, for the indicator of the set M we obtain that

$$p_e \le \sum_{x \ne x_0} \left(d_M(x_0 - x)\right)^t \le {d_M}^{t-1} \sum_{x \ne x_0} d_M(x_0 - x) =$$
$$= {d_M}^{t-1} |M|^{-1} \sum_{a \ne 0} \sum_{z \in G} I_M(z) I_M(z + a) =$$
$$= {d_M}^{t-1} |M|^{-1} \sum_{z \in G} I_M(z) \sum_{a \ne 0} I_M(z + a) =$$
$$= {d_M}^{t-1}(|M| - 1) \le |M| {d_M}^{t-1}.$$

Thus inequality (6) and the statement are proved.

Corollary 1. *Let condition (3) holds and* $d_M = 2^{c-(m-l)}$, *where* $c = \text{const}$. *Then the solution*

x_0 of system (2) can be found with probability at least $1-\delta$ in time $O\left(2^{m-l}\left(1+\frac{\log\delta^{-1}}{m-l}\right)\right)$ as $\delta\to 0$ and $m-l\to\infty$.

Note that in the worst case Algorithm 1 requires searching all $(m-l)$-dimensional Boolean vectors. Hence it becomes impractical, e.g., as $m-l\geq 64$.

3.2 Another Variants of the Attack in a Particular Case

Let's consider an important particular case, when in system (2)

$$(G,+)=(V_m,\oplus),\quad x_0=f_i(k),$$

$$M=\{(u,\phi(u)):\ u\in V_{m-l}\}. \tag{7}$$

Note that in this case parameter (5) coincides with the quantity

$$D_\phi=\max_{\alpha\in V_{m-l}\setminus\{0\},\,\beta\in V_l}\{2^{-(m-l)}\times$$

$$\times|\{z\in V_{m-l}:\phi(z\oplus\alpha)\oplus\phi(z)=\beta\}|\}, \tag{8}$$

which measures the resistance of the mapping ϕ against differential cryptanalysis (see [11], for example).

In this case, to solve system (2) we can apply another technique, which is sometimes more effective than Algorithm 1. This technique is related to linear cryptanalysis and correlation attacks on randomized stream ciphers with linear random coding [3 – 5, 7].

For any $n\in\mathbf{N}$, $a=(a_1,\dots,a_n)$, $b=(b_1,\dots,b_n)\in V_n$ denote $ab=a_1b_1\oplus\cdots\oplus a_nb_n$. Let's define

$$l_\phi(a,b)=2^{-(m-l)}|\{z\in V_{m-l}:az\neq b\phi(z)|,$$

$$a\in V_{m-l},\quad b\in V_l;$$

$$L_\phi=\max_{a\in V_{m-l},\,b\in V_l\setminus\{0\}}\{|1-2l_\varphi(a,b))|\}. \tag{9}$$

Consider the following algorithm of recovering the solution x_0 of system (2) under condition (7).

Algorithm 2.

1. Choose m linearly independent vectors (a_r,b_r), where $a_r\in V_{m-l}$, $b_r\in V_l\setminus\{0\}$ such that $l_\phi(a_r,b_r)\neq 1/2$, $r=\overline{1,m}$.

2. For every $r=\overline{1,m}$ obtain from (2) the system of equations

$$(a_ru_j\oplus b_r\phi(u_j))\oplus(a_r,b_r)x_0=(a_r,b_r)y_j,$$

$$j=\overline{1,t} \tag{10}$$

and recover the quantity $(a_r,b_r)x_0$ using the majority rule:

if $l_\phi(a_r,b_r)<1/2$, then

$$(a_r,b_r)x_0\overset{\text{def}}{=}0\ \Leftrightarrow\ \sum_{j=1}^{t}(a_r,b_r)y_j<t/2;$$

if $l_\phi(a_r,b_r)>1/2$, then

$$(a_r,b_r)x_0\overset{\text{def}}{=}0\ \Leftrightarrow\ \sum_{j=1}^{t}(a_r,b_r)y_j>t/2;$$

3. Find x_0 from the Obtained Quantities $(a_r,b_r)x_0$, using Gaussian Elimination

Note that step 1 and the transformation of the matrix with the rows (a_r,b_r), $r=\overline{1,m}$, on step 3 of Algorithm 2 are executed only once (at the stage of precomputation). Therefore the time complexity of this algorithm is determined by the execution time of step 2.

The proof of the following statement is almost the same as the proof of Statement 4 in [7].

Statement 2. *Under condition (7) the adversary can recover on step 2 of Algorithm 2 all values $(a_r,b_r)x_0$, $r=\overline{1,m}$, with probability at least $1-\delta$, $\delta\in(0,1)$, in $O(mt\log t)$ bit operations from*

$$t=\left\lceil 1/2\cdot\max_{1\leq r\leq m}\{|1-2l_\phi((a_r,b_r))|^{-2}\}\ln(\delta^{-1}m)\right\rceil$$

arbitrary equations of system (2).

Note that the data complexity, i.e., the number of equations in system (10), necessary for recovering one arbitrary quantity $(a_r,b_r)x_0$ on step 2 with probability at least $1-\delta$ is lower bounded by $C|1-2l_\phi((a_r,b_r))|^{-2}$, where the value C depends

only from δ. Therefore Algorithm 2 becomes impractical if the value of (9) is sufficiently small (e.g., $L_\phi \le 2^{-32}$).

In conclusion, consider another possible approach (based on ideas from algebraic cryptanalysis) for solving system (2) under condition (7).

Denote $x = (x_1, x_2)$, $\xi_j = (u_j, \phi(u_j))$, $y_j = (\alpha_j, \beta_j)$, where $x_1, u_j, \alpha_j \in V_{m-l}$, $x_2, \beta_j \in V_l$, $j = \overline{1,t}$. Then (2) is equivalent to the system of equations:

$$x_1 \oplus u_j = \alpha_j, \quad x_2 \oplus \phi(u_j) = \beta_j, \quad j = \overline{1,t},$$

which can be written as follows:

$$\phi(z \oplus (\alpha_1 \oplus \alpha_j)) \oplus \phi(z) = \beta_1 \oplus \beta_j, \quad j = \overline{2,t}, \quad (11)$$

$$x_1 = \alpha_1 \oplus z, \quad x_2 = \phi(z) \oplus \beta_1,$$

$$u_j = \alpha_j \oplus \alpha_1 \oplus z, \quad j = \overline{1,t}.$$

The following statement is oblivious.

Statement 3. *Under condition (7) the computational security of the considered cipher is upper bounded by the time complexity of solving the system of equations (11) for arbitrary (known) vectors* α_j, β_j, $j = \overline{1,t}$.

There a lot of families of Boolean mappings with small values of parameters (8) and (9) (see [11, 12], for example). But not all of them guarantee high complexity of solving systems of the form (11).

As an example, consider the mapping $\phi(x) = x^{2^l - 2}$, $x \in \mathbf{GF}(2^l)$, where $m = 2l$, l is even, widely used in modern block ciphers. It is known that $D_\phi = L_\phi{}^2 = 2^{2-l}$, but each separate equation of system (11) has at most four solutions, which can be found in real time [13]. Thus, having a small (namely, $t = \left\lceil \frac{l + \log \delta^{-1}}{\log(D_\phi^{-1})} \right\rceil + 1 = \left\lceil \frac{l + \log \delta^{-1}}{l-2} \right\rceil + 1$) number of equations in system (2), we can find in real time its (unique with probability at least $1 - \delta$, $\delta \in (0,1)$) solution x_0 by solving system (11).

At the same time, the problem of solving a system of the form (11) for arbitrary mapping $\phi : V_{m-l} \to V_l$ is computationally hard. Moreover, no efficient algorithms of solving such systems for any modern computationally secure hash functions are known so far. It seems very likely that the existence of such algorithms can be an undesirable property, which will allow us to distinct a hash function from the truly random mapping.

4. Conclusion

In contrast to before known approaches [1 – 5, 8, 9], the described framework gives more possibilities for designing computationally secure randomized stream ciphers. This is achieved by enlarging the class of transformations used in the construction of a randomizer. As the operation $*$ in (1) the addition modulo 2^m or the bitwise Boolean addition of binary vectors can be used, besides in the last case the matrix P can be chosen as the identity matrix of order m.

Under condition (7) the computational security of the proposed randomized stream ciphers in the chosen-plaintext-chosen-IV attacking scenario is determined by the following properties of the mapping $\phi : V_{m-l} \to V_l$:

(a) large value of $m - l$ to resist exhaustive search (Algorithm 1);

(b) small value of (9) to resist the linear-type attack (Algorithm 2);

(c) high time complexity of solving systems of equations (11).

In order to increase the practicality of the encryption scheme it is also desirable to choose the quantity l sufficiently large in comparison with m. For example, we can put $m - l = l = 128$ that provides encryption rate $l/m = 1/2$ independently from the choice of ϕ.

In order to resist the considered above (as well as other possible) attacks it is desirable that mapping ϕ

should have properties similar to those of random equiprobable mapping from V_{m-l} to V_l. From this point of view, it is natural to choose as ϕ one of modern keyless hash functions. We conjecture that any general attack on the randomized stream cipher can by efficiently transformed in an appropriate attack on the underlined hash function, but currently we don't have an accurate proof of this statement.

References

[1] M.J. Mihaljević, H. Imai, A stream ciphering approach based on wiretap channel coding, 8th Central European Conference of Cryptography 2008, Graz, Austria, July 2-4, E-Proc. (3 p.), 2008.

[2] M.J. Mihaljević, H. Imai, An approach for stream cipher design based on joint computing over random and secret data, Computing, 2009, Vol. 85, № 1-2, pp. 153-168.

[3] M.J. Mihaljević, H. Imai, An information-theoretic and computational complexity security analysis of a randomized stream cipher model, 4th Western European Workshop on Research in Cryptology, WeWoRC 2011, Weimar, Germany, July 20-22, Conf. Record, 2011, pp. 21-25.

[4] M.J. Mihaljević, H. Imai, Employment of homophonic coding for improvement of certain encryption approaches based on the LPN problem, Symmetric Key Encryption Workshop, SKEW 2011, Copenhagen, Denmark, Feb. 16-17, E-Proc. (17 p.), 2011.

[5] M.J. Mihaljević, F. Oggier, H. Imai, Homophonic coding design for communication systems employing the encoding-encryption paradigm, arXiv:1012.5895v1 [cs.CR], 29 Dec, 2010.

[6] E.R. Berlekamp, R.J. McElice, H. van Tilborg, On the inherent intractability of certain coding problems, IEEE Trans. on Inform. Theory, 1978, Vol. 24, No. 3, pp. 384-386.

[7] A.N. Alekseychuk, S.V. Gryshakov, On the computational security of randomized stream ciphers proposed by Mihaljević and Imai, Zakhist Inform., 2014, No. 4, pp. 328-334

[8] A.N. Alekseychuk, Analytical estimates of theoretical security of randomized block ciphers against differential cryptanalysis, Zakhist Inform., 2007, No. 3, pp. 80-88 (in Russian).

[9] A.N. Alekseychuk, Sufficient conditions for randomized block cipher-systems to be secure against commutative diagram cryptanalysis, Data Recording, Storage and Processing, 2007, Vol. 9, No. 2, pp. 61-68, (in Russian).

[10] ECRYPT II: Final hash function status report, http://www.ecrypt.eu.org/documents/D.SYM.11, 31 Jan., 2013.

[11] A. Canteaut, Cryptographic functions and design criteria for block ciphers, INDOCRYPT 2001, LNCS 2247, Springer Verlag, 2001, pp. 1-16.

[12] C. Carlet, Vectorial Boolean functions for cryptography, in "Boolean Models and Methods in Mathematics, Computer Science and Engeneering", Cambridge University Press, 2010, pp. 398-469.

[13] K. Nyberg, Differentially uniform mappings for cryptography, Advances in Cryptology, EUROCRYPT'93, LNCS 765, Springer Verlag, 1994, pp. 55-64.