UDC 004.056.55

PERIODIC CHARACTERISTICS OF OUTPUT FEEDBACK ENCRYPTION MODE

Alexandr Kuznetsov, Ievgeniia Kolovanova, Tetiana Kuznetsova

V. N. Karazin Kharkiv National University, Svobody sq., 4, Kharkov, 61022, Ukraine
kuznetsov@karazin.ua, e.kolovanova@gmail.com, kuznetsova.tatiana17@gmail.com

Reviewer: Serhii Toliupa, Doctor of Sciences (Engineering), Full Prof., Taras Shevchenko National University of Kiev, Lomonosova St., 81, Kyiv, 03189, Ukraine.
tolupa@i.ua

Received on May 2017

Abstract. *We investigate periodic characteristics of sequence of output blocks in the output feedback encryption mode. The model of random homogeneous substitution is used for an abstract description of this formation. This property is directly related to the periodic properties of output feedback encryption mode, since it characterizes the probabilistic distribution of output blocks with certain period appearance, provided that the assumption is made that the properties of the block symmetric cipher are consistent with certain properties of the random substitution. Also in the work specific practical tasks are solved, namely recommendations are being developed for the application of the outbound feedback on the encryption threshold, certain requirements and limitations are justified.*

Keywords: *encryption mode; random substitution; periodic characteristics of output blocks; output feedback.*

1 Introduction

One of the most common block symmetric encryption modes used to provide confidentiality services is Outbound Feedback (OFB). The OFB has several advantages: firstly, the output block can be formed in advance, even before the message, which can greatly speed up the process of protecting information; and secondly, in this mode, as in the mode Electronic Codebook - ECB, errors that can arise when transmitting ciphertext over the communication channels, are localized in the block, not extending to the neighboring, and in the OFB mode only changed bits will be false (in the ECB mode the entire block will change). Thirdly, the cryptographic properties of the output block do not depend on the open text, they are determined only by the properties of the base cryptographic transformation, and, possibly, by the value of the initialization block, which determines the specific form and frequency of the output block. This work is devoted to study of the periodic properties of output block in OFB mode, because the occurrence of output block repetition is the most dangerous case, it gives an attacker the possibility to violate the established mode of message confidentiality.

We derive a formula for estimating the probability of a cycle occurrence for an arbitrary fixed value of a set of transformations. This property is directly related to the periodic properties of the output blocks, since it characterizes the probabilistic distribution of the output block with certain period appearance, provided that the assumption is made that the properties of the block symmetric cipher are consistent with certain properties of the random substitution. Also in the work specific practical tasks are solved, namely recommendations are being developed for the application of the Outbound Feedback on the encryption threshold, certain requirements and limitations are justified. The conclusions summarize and concretize our results, discuss possible ways of further research.

2. Output Feedback Encryption Mode

The output feedback encryption mode is intended to provide a confidentiality service. This mode is based on encryption of the initialization vector to generate a sequence of output blocks that are added to the normal text to form encrypted text and, conversely, to the ciphertext to decrypt it. This mode requires unique initialization vector for each application with the provided (fixed) key. Let's

consider the specification of output feedback encryption mode.

The parameters of the mode are encryption key K, $|K|=k$, and initialization vector S, $|S|=l$. Additional requirements for initialization vector are not imposed. When encrypting the message M ($|M|\geq 1$) is presented as a sequence of blocks:

$$M=m_1\|m_2\|...\|m_n,\ |m_i|=l$$

for $i=1,2,..,n-1$, $1\leq|m_n|\leq l$.

The initial value of the output block γ_0 ($|\gamma_0|=l$) is calculated as

$$\gamma_0=T_{l,k}^{(K)}(S). \quad (1)$$

Each ciphertext block is calculated according to the ratio

$$c_i=m_i\oplus L_{l,|m_i|}(\gamma_{i-1}) \quad (2)$$

for $i=1,2,..,n$, and

$$\gamma_i=T_{l,k}^{(K)}(\gamma_{i-1}) \quad (3)$$

for $i=1,2,..,n-1$.

The result of message encrypting is ciphertext $C=c_1\|c_2\|...\|c_n$.

When decrypting, the ciphertext C ($|C|\geq 1$) is presented in the form of a sequence of blocks:

$$C=c_1\|c_2\|...\|c_n,\ |c_i|=l$$

for $i=1,2,..,n-1$, $1\leq|c_n|\leq l$.

The initial value of the output block γ_0 ($|\gamma_0|=l$) is calculated as

$$\gamma_0=T_{l,k}^{(K)}(S).$$

Each message block is calculated according to the ratio

$$m_i=c_i\oplus L_{l,|m_i|}(\gamma_{i-1}) \quad (4)$$

for $i=1,2,..,n$, and

$$\gamma_i=T_{l,k}^{(K)}(\gamma_{i-1})$$

for $i=1,2,..,n-1$.

The result of the decryption of ciphertext is the message $M=m_1\|m_2\|...\|m_n$.

The encryption and decryption scheme in the output feedback encryption mode is shown in Fig. 1. This scheme formally depicts the sequence of execution of transformations (2), (4) for all values of the cyclic variable $i=1,2,..,n$.

Let's consider the periodic characteristics of output feedback encryption mode. First, we note that, by definition, the output block consists of the initial value of the block γ_0 and the remaining blocks γ_i, which are calculated for (1), (3) for each value of the cyclic variable $i=1,2,..,n-1$. That is, the task of the research is precisely in determining the period of the sequence of blocks $\gamma_i, i=0,1,..,n-1$.

Each output block γ_i is the result of encrypting the previous block γ_{i-1}, where the initial value γ_0 is equal to the result of the initialization vector encryption. If you use the terminology of the substitutions theory [1-3] and present the basic encryption transformation $T_{l,k}^{(K)}$ initiated by the secret key K as some substitution s acting on the set of open texts, then the period of the sequence of

output blocks γ_0, γ_1, …, γ_{n-1} will correspond to one of the cycles $s_i = (y, s_i(y), s_i^2(y), ..., s_i^{l_i-1}(y))$ of the substitution s where the initial value of the cycle y is equal to the initialization vector value S.

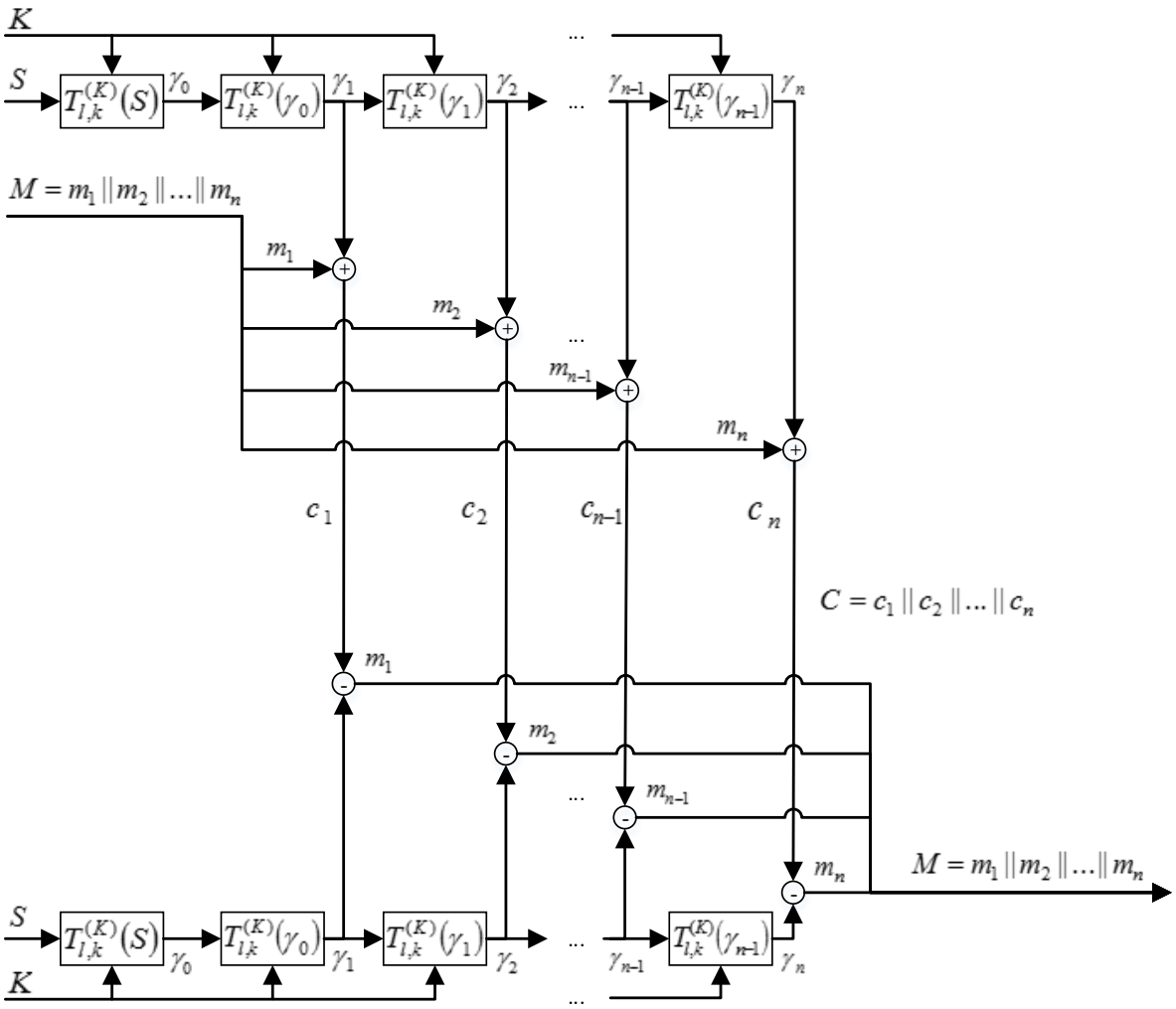

Fig. 1 – Scheme of encryption and decryption

Each next element $s_i^j(y)$ of the cycle s_i with the length l_i is the result of multiple encoding of the initialization vector:

$$y = S;$$

$$s_i(y) = T_{l,k}^{(K)}(S) = \gamma_0;$$

$$s_i^2(y) = T_{l,k}^{(K)}\left(T_{l,k}^{(K)}(S)\right) = \gamma_1;$$

$$\dots$$

$$s_i^{l_i-1}(y) = \underbrace{T_{l,k}^{(K)}\left(T_{l,k}^{(K)} \dots T_{l,k}^{(K)}(S)\right)}_{l_i-1 \text{ разів}} = \gamma_{l_i-2};$$

$$s_i^{l_i}(y) = \underbrace{T_{l,k}^{(K)}\left(T_{l,k}^{(K)} \dots T_{l,k}^{(K)}(S)\right)}_{l_i \text{ разів}} = \gamma_{l_i-1} = \gamma_0;$$

$$\dots$$

$$s_i^n(y) = \underbrace{T_{l,k}^{(K)}\left(T_{l,k}^{(K)} \dots T_{l,k}^{(K)}(S)\right)}_{n \text{ разів}} = \gamma_{n-1}.$$

Thus, the investigation of the periodic properties of the sequence of output blocks $\gamma_i, i = 0,1,..,n$ is to study the cyclic structure of the substitution s, namely, in estimating the distribution of the length l_i of the cycles s_i for different initial values $y = S$ of the basic ciphering transformation $T_{l,k}^{(K)}$. Such investigations will make it possible to estimate the length l_i of output blocks period for different $y = S$ and K or to determine the probability of forming the output blocks of a certain period for an arbitrary fixed value of the initialization vector $y = S$ and a randomly selected secret key K.

Let us consider some of the provisions of the theory of substitutions and their relation to the properties of BSC, in particular, we introduce the basic concepts and definitions associated with certain properties of symmetric block crypto-transformations (the distribution of the number of cycles, magnifications and inversions, etc.).

3. Special Provisions of the Theory of Substitutions

By the definition BSC is a key-parameterized function of a bijective mapping of a set of plaintexts into a set of ciphertexts $V_l \to V_l$, $K \in V_k$ [4]. In general, for any *l*-bit block cipher there are $2^l!$ possible permutations of plaintext. These transformations, called permutations of degree 2^l, form a group under the operation of performing sequential transformations. Such a group is called a symmetric group of permutations of degree 2^n and is denoted by S_{2^l} [1]. In practice, it means that the number of bits of the key, which is necessary to obtain all possible permutations, is about $\ln 2^l! \approx l \cdot 2^l$ bits (by the Stirling formula $\ln(x!) = x\ln(x) - x - O(\ln(x))$). For example, for l = 128, we have $2^{128}! \approx 2^{128 \cdot 2^{128}}$ possible permutations of 128-bits blocks, of which, depending on the length of the key, only 2^{128} or 2^{256} transformations are used. Thus, the basic transformation of the cipher is essentially a subset of the complete set of all possible substitutions acting on a set of processed data blocks. The basic assumption that is adopted in substantiating the stability of symmetric cryptographic transformation is in preserving the probabilistic properties of random substitution. It is assumed that while encrypting and applying a limited set of substitutions from S_{2^l}, however, certain distribution probabilities of this subset elements correspond to the properties of randomly selected substitutions from the whole set S_{2^l} [5-8].

Let's consider the basic concepts and definitions of the theory of substitutions [1] and associate them with cyclic properties of BSC. For this we consider the set of all biective transformations of the set $Y = \{y_1, y_2, ..., y_n\}$ to itself, forming a symmetric group S_n with the power $n!$ of all possible substitutions of degree *n*. By definition of the symmetric group [1], each substitution $s \in S_n$ corresponds to a unique substitution $s^{-1} \in S_n$, such that

$$s^{-1} \cdot s(y) = s \cdot s^{-1}(y) = e(y), \; y \in Y,$$

where $e(y) \in S_n$ is the unit substitution, i.e. $e(y) = y$ for all $y \in Y$.

Let's use the following symbols:

$$s \cdot s \cdot ... \cdot s = s^k, \; s^{-1} \cdot s^{-1} \cdot ... \cdot s^{-1} = s^{-k},$$

where products contain k multipliers. Accordingly, we have

$$s^k \cdot s^{-k} = s^{-k} \cdot s^k = s^0 = e.$$

The set of substitutions of degree *n*, which is locked in relation to the multiplication and inverse computation operation for $s \in S_n$ an element $s^{-1} \in S_n$, is called substitution group. Each such group is a subgroup of the symmetric group S_n [1].

Consider some substitutions $s \in S_n$ that act on a set Y. We will define a binary relation on a set Y, while we will assume that $y \sim y'$ for $y, y' \in Y$ if there exists such j that $y' = s^j(y)$. This bina-

ry relation is reflexive, symmetric and transitive, i.e. the relation is equivalence. Indeed, according to [1] we have:

- $y \sim y$, because $y = s^0(y) = e(y)$;
- - it follows from condition $y \sim y'$ that $y' \sim y$, because it follows from equality $y' = s^j(y)$ that $y = s^{-j}(y')$;
- - it follows from $y \sim y'$ and $y' \sim y''$ that $y \sim y''$, because from the equalities $y' = s^j(y)$ and $y'' = s^i(y')$ it follows that $y'' = s^i(s^j(y)) = s^{i+j}(y)$.

The cycle s_i of substitution $s \in S_n$ with the length l_i is defined as follows:

$$s_i = (y, s_i(y), s_i^2(y), ..., s_i^{l_i-1}(y)),$$

where $s_i^{l_i}(y) = y$.

An arbitrary substitution $s \in S_n$ can be expanded into the corresponding cycles [1]:

$$s = (y_1, s_1(y_1), s_1^2(y_1), ..., s_1^{l_1-1}(y_1))...(y_k, s_k(y_k), s_k^2(y_k), ..., s_k^{l_k-1}(y_k)). \quad (5)$$

Elements y_i and y_{i+1} in substitution $s \in S_n$ form increment, if $s(y_i) > s(y_{i+1})$ it is assumed that an element y_1 always preceded by increment. A pair of elements y_i and y_j in substitution $s \in S_n$ forms an increment if $s(y_i) > s(y_j)$, $i < j$.

For example, the substitution s of degree 4

$$s = \begin{pmatrix} y_1 & y_2 & y_3 & y_4 \\ s(y_1) & s(y_2) & s(y_3) & s(y_4) \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}$$

can be given in the form of a schedule for 3 cycles:

$$s_1 = (y_1) = (1),\ l_1 = 1;$$
$$s_2 = (y_2, s_2(y_2)) = (2,4),\ l_2 = 2;$$
$$s_3 = (y_3) = (3),\ l_3 = 1.$$

We have the following schedule:

$$s = (y_1)(y_2, s_2(y_y))(y_3) = (1)(2,4)(3).$$

In this substitution there are two increments (the element $y_1 = 1$ always preceded by one increment, and one more increment forms the elements $y_1 = 1$ and $y_2 = 2$, because $s(y_1) = 1 > s(y_2) = 4$) and three inversions (they form pairs of elements y_2 and y_3, y_2 and y_4, y_3 and y_4, because there are inequalities $s(y_2) > s(y_3)$, $s(y_2) > s(y_4)$, $s(y_3) > s(y_4)$, respectively).

On the set of all permutations of the symmetric group S_n, we give a uniform probabilistic distribution, i.e. for each selected substitution $s \in S_n$ we put in correspondence the probability of its selection equal to $1/n!$. According to modern views of symmetric cryptography, such a set of equivalence mappings corresponds to the idea of an "ideal" cipher. After all, if the random selection of a separate substitution $s \in S_n$ is associated with the value of the entered encryption key, then the resulting conversion will match the random and evenly selected ciphertext for each open text with any key, i.e. in all possible variants of open text mapping in the ciphertext.

We will investigate the probabilistic properties of random substitution, in particular, the probabilities of a cycle of a certain length in a randomly selected substitution, since this particular event will correspond to the case when the output block γ_i of a certain period is formed for an arbitrary fixed value of the initialization vector S.

4. Probability Evaluation of a Certain Length Cycle in Randomly Selected Substitution

Consider a random value ξ_n equals to the number of cycles in a randomly chosen substitution $s \in S_n$. The substitution $s \in S_n$ refers to a cyclic class $\{1^{\alpha_1}2^{\alpha_2}...n^{\alpha_n}\}$ if it contains α_1 cycles of length 1, α_2 cycles of length 2, and so on:

$$s = (y_1)(y_2)...(y_{\alpha_1})(y'_1, y''_1)(y'_2, y''_2)...(y'_{\alpha_2}, y''_{\alpha_2})...,$$
$$1\alpha_1 + 2\alpha_2 + ... + n\alpha_n = n.$$

Denote by $C(\alpha_1, \alpha_2, ..., \alpha_n)$ the number of substitutions in the cyclic class $\{1^{\alpha_1}2^{\alpha_2}...n^{\alpha_n}\}$, and by $C(n,k)$ the number of substitutions of degree n that have k cycles. Then we have [1]:

$$C(\alpha_1, \alpha_2, ..., \alpha_n) = \frac{n!}{1^{\alpha_1}2^{\alpha_2}...n^{\alpha_n}\alpha_1!\alpha_2!...\alpha_n!},$$

$$C(n,k) = \sum_{\substack{1\alpha_1+2\alpha_2+...+n\alpha_n=n \\ \alpha_1+\alpha_2+...+\alpha_n=k,\ \alpha_i\geq 0}} \frac{n!}{1^{\alpha_1}2^{\alpha_2}...n^{\alpha_n}\alpha_1!\alpha_2!...\alpha_n!} = |s(n,k)|, \quad (6)$$

where $s(n,k)$ are Stirling numbers of first kind, which is determined by the ratio

$$(x)_n = x(x-1)...(x-n+1) = \sum_{k=0}^{n} s(n,k)x^k,$$

where $(x)_n = x(x-1)...(x-n+1)$ is the common designation of the declining factorial (*the symbol of Pohgammer*).

Formula (6) implies the expression for the exact probability distribution of a random event $\xi_n = k$, in the case where randomly selected substitutions will observe exactly k cycles (see expression (5)).

Using the formula for computing Stirling numbers of the first kind, we have [1]:

$$P(\xi_n = k) = \frac{C(n,k)}{n!} = \frac{|s(n,k)|}{n!},\ k = 0,1,...,n.$$

In [1] we obtain the expected value $M\xi_n$ and variance $D\xi_n$ of the random variable ξ_n:

$$M\xi_n = \sum_{j=1}^{n}\frac{1}{j} = \ln n + C + o(1),\ D\xi_n = \sum_{j=1}^{n}\frac{1}{j} - \sum_{j=1}^{n}\frac{1}{j^2} = \ln n + C + o(1),\ C = 0,5772...,$$

in addition, it is shown that when $n \to \infty$ a random variable $\xi'_n = (\xi_n - \ln n)/(\ln n)$ is distributed asymptotically with parameters $(0,\ 1)$

$$\lim_{n\to\infty} P(\xi'_n < u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} e^{-y^2/2}dy$$

For random variables ζ_n and η_n, which are equal to the number of increments and inversions in randomly selected substitution $s \in S_n$, the corresponding expected value and variance have the form [1]:

$$M\zeta_n = \frac{n(n-1)}{4},\ D\zeta_n = \frac{2n^3+3n^2-5n}{72},\ M\eta_n = \frac{n}{2},\ D\eta_n = \frac{n}{12},$$

in this case random variables

$$\zeta'_n = (\zeta_n - M\zeta_n)/(D\zeta_n)$$

when $n \to \infty$ are also asymptotically distributed with parameters $(0,\ 1)$.

Empirical distributions of the probability of occurrence of a certain number of cycles, increments and inversions in randomly selected substitutions from a certain subset $V \subset S_n$, whose elements are substitutions implemented by the use of the encryption function on reduced cipher models, are investigated in [5, 8]. It is established that the obtained empirical distributions are very close to the theoretical distributions under consideration, i.e. it can be argued that the reduced models of BSC on these criteria are similar to the properties of random substitution from S_n.

At the same time, to evaluate the probability of forming output blocks γ_i of a certain period for an arbitrary fixed value of the initialization vector S another characteristic of random substitution is required, namely, the distribution of the number of cycles of a given length. In accordance with [1], this characteristic in random substitution is determined as follows.

We denote $\chi_{n,L}$ as the number of cycles of length L in an arbitrary equivalence arbitrary substitution of degree *n*. Obviously that

$$\xi_n = \chi_{n,L=1} + \chi_{n,L=2} + ... + \chi_{n,L=n}.$$

The probability distribution of a random event $\chi_{n,L} = k$ is defined as [1]:

$$P(\chi_{n,L} = k) = \frac{1}{L^k k!} \sum_{j=0}^{[n/L]-k} \frac{(-1)^j}{L^j j!},\ k = 0,1,...,\lfloor n/L \rfloor. \quad (7)$$

When $n \to \infty$ a random variable $\chi_{n,L}$ has a Poisson distribution with parameters $\lambda = 1/L$, i.e.

$$\lim_{n\to\infty} P(\chi_{n,L} = k) = \frac{1}{L^k k!} e^{-1/L},\ k = 0,1,.... \quad (8)$$

We use the formula for the exact distribution of probabilities of a random event $\chi_{n,L} = k$ in the form (7) [1]. The value $n!P(\chi_{n,L} = k)$ corresponds to the number of substitutions containing k cycles with the length L. We are interested in the number of such substitutions $s \in S_n$, which for an arbitrary fixed $y_i \in Y$ will necessarily have cycles $s_i = (y_i, s_i(y_i), s_i^2(y_i), ..., s_i^{l_i-1}(y_i))$ of lengths $L = l_i$. Consider the case when $L = 1$, i.e. we will count the number of such substitutions from S_n which, for an arbitrary fixed $y_i \in Y$ will necessarily have a cycle (y_i) of length $L = l_i = 1$. In cryptography, when considering block symmetric cryptographic transformations, such cases are called fixed points of substitution [5]. Taking into account that $L = 1$, formula (7) takes the form

$$P(\chi_{n,L=1} = k) = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!},\ k = 0,1,...,n,$$

and for each $k = 1,...,n$ of $n!P(\chi_{n,L=1} = k)$ cases for an arbitrary fixed $y_i \in Y$ will be observed precisely

$$\frac{C_{n-1}^{k-1}}{C_n^k} = \frac{(n-1)!}{(k-1)!(n-k)!} \frac{k!(n-k)!}{n!} = \frac{k}{n} \quad (9)$$

times, i.e. the number of substitutions containing one fixed point of a specific form (cycle (y_i)) will be determined by the formula:

$$\sum_{k=1}^{n} n!P(\chi_{n,L=1} = k) \frac{C_{n-1}^{k-1}}{C_n^k} = \sum_{k=1}^{n} \left(\frac{(n-1)!}{(k-1)!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \right) = (n-1)!, \quad (10)$$

and the corresponding probability of the appearance of such fixed point (for given initial value $y_i \in Y$) in the randomly chosen substitution of the degree n will look like:

$$\sum_{k=1}^{n} P(\chi_{n,L=1} = k) \frac{C_{n-1}^{k-1}}{C_n^k} = \frac{(n-1)!}{n!} = \frac{1}{n}. \quad (11)$$

Let's explain the formula (9). In total, there are exactly C_n^k methods for simultaneously choosing values $y_i, y_{i_1}, y_{i_2}, ..., y_{i_{k-1}} \in Y$, $y_i \neq y_{i_1} \neq y_{i_2} \neq ... \neq y_{i_{k-1}}$ which uniquely determine cycles $(y_i)(y_{i_1})(y_{i_2})...(y_{i_{k-1}})$ of length $L=1$. But for each fixed $y_i \in Y$ there are exactly C_{n-1}^{k-1} ways for choosing the remaining values $y_{i_1}, y_{i_2}, ..., y_{i_{k-1}} \in Y$. I.e. from the total number $n!P(\chi_{n,L=1}=k)$ of substitutions containing *k* cycles of length $L=1$, only

$$n!P(\chi_{n,L=1}=k)\frac{C_{n-1}^{k-1}}{C_n^k} = n!P(\chi_{n,L=1}=k)\frac{k}{n}$$

substitutions will necessarily contain a cycle (y_i).

The last formula (11) can be obtained much simpler from trivial combinatorial considerations. Indeed, if on the set $Y=\{y_1, y_2, ..., y_n\}$ we will fix m elements, then there are $(n-m)!$ ways for permutations of the remaining elements. I.e. on the all set of substitutions from S_n with the random probability distribution, the probability of choosing a substitution with m fixed points is equal to

$$\frac{(n-m)!}{n!} = \frac{1}{(n-m+1)(n-m+2)...n} = \frac{1}{(n)_m}, \quad (12)$$

Which for $m=1$ coincides with (11).

Formulas (9-12) were obtained in [9] when studying Galois / Counter Mode and GMAC - GCM & GMAC, which allowed us to estimate the probability of a zero hash subkey, i.e. the probability of such event, when encryption of zero open text will get zero value of ciphertext. We extend the result obtained earlier to an arbitrary value of the length of cycle $L=l_i \in \{1,2,...,n\}$ in the study of periodic properties of the output blocks in OFB mode.

Consider the case of an arbitrary length of a cycle, we will calculate the number of such substitutions s from S_n which, for an arbitrary fixed $y_i \in Y$ will necessarily have a cycle

$$(y_i, s_i(y_i), s_i^2(y_i), ..., s_i^{L-1}(y_i))$$

of length $L=l_i \in \{1,2,...,n\}$.

For fixed lengths and quantities of cycles (L and k) there are exactly C_n^{kL} ways to simultaneously select values

$$\begin{gathered} y_i, y_{j_i}, y_{u_i}, ..., y_{v_i}, \\ y_{i_1}, y_{j_1}, y_{u_1}, ..., y_{v_1}, \\ y_{i_2}, y_{j_2}, y_{u_2}, ..., y_{v_2}, \\ ..., \\ y_{i_{k-1}}, y_{j_{k-1}}, y_{u_{k-1}}, ..., y_{v_{k-1}}, \end{gathered} \quad (13)$$

which collectively determine k cycles of length L each:

$$\begin{gathered} (y_i, y_j = s_i(y_i), y_u = s_i^2(y_i), ..., y_v = s_i^{L-1}(y_i)), \\ (y_{i_1}, y_{j_1} = s_{i_1}(y_{i_1}), y_{u_1} = s_{i_1}^2(y_{i_1}), ..., y_{v_1} = s_{i_1}^{L-1}(y_{i_1})), \\ (y_{i_2}, y_{j_2} = s_{i_2}(y_{i_2}), y_{u_2} = s_{i_2}^2(y_{i_2}), ..., y_{v_2} = s_{i_2}^{L-1}(y_{i_2})), \\ ..., \\ (y_{i_{k-1}}, y_{j_{k-1}} = s_{i_{k-1}}(y_{i_{k-1}}), y_{u_{k-1}} = s_{i_{k-1}}^2(y_{i_{k-1}}), ..., y_{v_{k-1}} = s_{i_{k-1}}^{L-1}(y_{i_{k-1}})), \end{gathered} \quad (14)$$

and all elements from (13) are unique, since the set of cycles (14) is included in the decomposition of the same substitution.

Of the C_n^{kL} ways of simultaneously choosing the values (13) for each fixed set $y_i, y_{j_i}, y_{u_i}, ..., y_{v_i} \in Y$ there are exactly C_{n-1}^{kL-1} ways for choosing the remaining values

$$y_{i_1}, y_{j_1}, y_{u_1}, ..., y_{v_1},$$
$$y_{i_2}, y_{j_2}, y_{u_2}, ..., y_{v_2},$$
$$...,$$
$$y_{i_{k-1}}, y_{j_{k-1}}, y_{u_{k-1}}, ..., y_{v_{k-1}},$$

because the selection of set $y_i, y_{j_i}, y_{u_i}, ..., y_{v_i} \in Y$ is determined by selecting only one element $y_i \in Y$, and from the remaining $n-1$ elements possible different combinations of $kL-1$ elements.

Thus, for each $k = 0,1,...,\lfloor n/L \rfloor$ of the total number $n!P(\chi_{n,l} = k)$ of substitutions containing k cycles of length L, only

$$n!P(\chi_{n,L} = k)\frac{C_{n-1}^{kL-1}}{C_n^{kL}} = n!P(\chi_{n,L} = k)\frac{(n-1)!kL!(n-kL)!}{n!(kL-1)!(n-kL)!} = n!P(\chi_{n,L} = k)\frac{kL}{n} = (n-1)!kLP(\chi_{n,L} = k)$$

substitutions will necessarily contain a cycle

$$(y_i, s_i(y_i), s_i^2(y_i), ..., s_i^{L-1}(y_i)).$$

Summing up the last expression for all $k = 0,1,...,\lfloor n/L \rfloor$, taking into account (7), we obtain an exact formula for determining the number of substitutions s from S_n which, for an arbitrary fixed $y_i \in Y$ will necessarily have a cycle

$$(y_i, s_i(y_i), s_i^2(y_i), ..., s_i^{L-1}(y_i))$$

of length $L = l_i \in \{1, 2, ..., n\}$:

$$\sum_{k=1}^{[n/L]} n!P(\chi_{n,L} = k)\frac{C_{n-1}^{kL-1}}{C_n^{kL}} = \sum_{k=1}^{[n/L]} \frac{(n-1)!}{(k-1)!L^{k-1}} \sum_{j=0}^{[n/L]-k} \frac{(-1)^j}{L^j j!} = (n-1)!, \qquad (15)$$

and the corresponding formula for calculating the probability of randomly choosing a substitution s from S_n with the following cycle:

$$\sum_{k=1}^{[n/L]} P(\chi_{n,L} = k)\frac{C_{n-1}^{kL-1}}{C_n^{kL}} = \frac{(n-1)!}{n!} = \frac{1}{n}. \qquad (16)$$

It is obvious that the last analytic expression for $L = 1$ completely coincides with the formula (8) in [9] with the corresponding statement.

The resulting analytic expression (16) can also be considered as a combinatorial identity (simplified formula) for the sum of the members of formula (7) with the corresponding proportional coefficients

$$\frac{C_{n-1}^{kL-1}}{C_n^{kL}} = \frac{kL}{n},$$

or even for the Poisson distribution (8). The probability estimate (16) of a cycle of a certain length can be obtained by another, in a much simpler way, using simple combinatorial considerations.

We fix some arbitrary value y_i from the set $Y = \{y_1, y_2, ..., y_n\}$. There are totally $n!$ substitutions on the set Y, of which only

$$\frac{n!}{n} = (n-1)!$$

substitutions will contain a cycle (y_i) of length $L = 1$ in their cyclic schedule.

In addition, from $n!$ substitutions of the symmetric group

$$\frac{n!}{n(n-1)}(n-1) = (n-1)!$$

substitutions will contain a cycle $(y_i, y_{j\neq i})$ of length $L=2$,

$$\frac{n!}{n(n-1)(n-2)}(n-1)(n-2)=(n-1)!$$

substitutions will contain a cycle $(y_i, y_{j\neq i}, y_{u\neq i,j})$ of length $L=3$ and so on.

That is, for an arbitrary fixed value $y_i \in Y$ the number of substitutions from S_n containing the cycle $(y_i, s_i(y_i), s_i^2(y_i), ..., s_i^{L-1}(y_i))$ is defined as $(n-1)!$, and the corresponding probability of randomly choosing a substitution containing such a cycle is defined as

$$\frac{(n-1)!}{n!}=\frac{1}{n},$$

regardless of the length of the cycle $L=l_i \in \{1,2,...,n\}$, nor its own value y_i from $Y=\{y_1, y_2, ..., y_n\}$.

Thus, the probability of a cycle of a certain length is determined only by the degree n of substitution. For example, for $n=4$ from $n!=24$ substitutions of the symmetric group for any fixed y_i from $Y=\{y_1, y_2, y_3, y_4\}$ we have $(n-1)!=6$ substitutions each that necessarily contain cycles of different lengths (or (y_i), or $(y_i, y_{j\neq i})$, or $(y_i, y_{j\neq i}, y_{u\neq i,j})$, or $(y_i, y_{j\neq i}, y_{u\neq i,j}, y_{v\neq i,j,u})$, respectively). Consequently, the probability that a randomly selected substitution from S_4 contained cycle of length $L=l_i \in \{1,2,...,4\}$ will be equal to $1/n=1/4$ independent from either y_i no $L=l_i$.

Since the obtained analytical expressions (15) and (16) are rather complicated and cumbersome, especially the order of their output, we illustrate the example of calculating the probabilities of a cycle of given length in a randomly chosen substitution from a symmetric group S_4. An example will be supplemented by explanations showing the validity of the formulas obtained and the combinatorial arguments presented.

5 Example For Symmetric Group S_4

Consider an example of all biective transformations of the set $Y=\{y_1, y_2, y_3, y_4\}$ to itself, i.e. the set with $n!=24$ substitutions of degree *n = 4*.

Table 1 shows all substitutions that make up the symmetric group S_4 (the results of each substitution are given, the order of each substitution for cycles, total number of cycles and the distribution of the number of cycles of a certain length, each substitution is numbered for convenience).

Table 2 shows the distribution of the number of values $\xi_n=k$ and $\chi_{n,L}=k$ for different k (the symbol # (x) indicates the number of cases *(x)* for all substitutions from S_4).

Consider the case when for an arbitrary fixed $y_i \in Y$ randomly chosen substitution *s* from S_4 will necessarily contain a cycle (y_i) of length $L=1$.

First we consider substitutions with *k = 1* cycles of length $L=1$. We have 8 such substitutions (table 2). But each individual $y_i \in Y$ generates a cycle of length $L=1$ only twice, that is, for each arbitrary fixed $y_i \in Y$ there are precisely 2 substitutions that contain a cycle of length $L=1$ of the form (y_i). For example, for $y_1 \in Y$ these are 4th and 5th substitutions, for $y_2 \in Y$ these are the 16th and 21st substitutions, etc. The number of substitutions, which for an arbitrary fixed $y_i \in Y$ containing k = 1 cycle of length $L=1$ of form (y_i) are calculated as follows. In total, there are exactly $C_{n=4}^{kL=1}=4$ ways to select a value $y_i \in Y$. This choice is no longer limited, because for each $y_i \in Y$ cycle (y_i) is defined unambiguously (according to the formula there are $C_{n-1=3}^{kL-1=0}=1$ options for choosing a value $y_{j\neq i} \in Y$). That is, the total number of substitutions containing only k = 1 cycle of

length $L=1$ (such 8 substitutions) must be multiplied by value $\frac{C_{n-1=3}^{kL-1=0}}{C_{n=4}^{kL=1}}=\frac{1}{4}$. Thus, for an arbitrary fixed $y_i \in Y$ the number of substitutions containing only one cycle (y_i) is equal to two.

Table 1 – The set of substitutions from S_4 and their cyclic properties

№	Result of substitution				Decomposition of substitution to cycles	Number of cycles, ξn	Number of cycles of length L, χn, L			
	$s(y_1)$	$s(y_2)$	$s(y_3)$	$s(y_4)$			L=1	L=2	L=3	L=4
1	y_1	y_2	y_3	y_4	$(y_1)(y_2)(y_3)(y_4)$	4	4	0	0	0
2	y_1	y_2	y_4	y_3	$(y_1)(y_2)(y_3, y_4)$	3	2	1	0	0
3	y_1	y_3	y_2	y_4	$(y_1)(y_2,y_3)(y_4)$	3	2	1	0	0
4	y_1	y_3	y_4	y_2	$(y_1)\ (y_2, y_3, y_4)$	2	1	0	1	0
5	y_1	y_4	y_2	y_3	$(y_1)\ (y_2, y_4, y_3)$	2	1	0	1	0
6	y_1	y_4	y_3	y_2	$(y_1)\ (y_2, y_4)\ (y_3)$	3	2	1	0	0
7	y_2	y_1	y_3	y_4	$(y_1, y_2)\ (y_3)\ (y_4)$	3	2	1	0	0
8	y_2	y_1	y_4	y_3	$(y_1, y_2)\ (y_3, y_4)$	2	0	2	0	0
9	y_2	y_3	y_1	y_4	$(y_1, y_2, y_3)\ (y_4)$	2	1	0	1	0
10	y_2	y_3	y_4	y_1	(y_1, y_2, y_3, y_4)	1	0	0	0	1
11	y_2	y_4	y_1	y_3	(y_1, y_2, y_4, y_3)	1	0	0	0	1
12	y_2	y_4	y_3	y_1	$(y_1, y_2, y_4)\ (y_3)$	2	1	0	1	0
13	y_3	y_1	y_2	y_4	$(y_1, y_3, y_2)\ (y_4)$	2	1	0	1	0
14	y_3	y_1	y_4	y_2	(y_1, y_3, y_4, y_2)	1	0	0	0	1
15	y_3	y_2	y_1	y_4	$(y_1, y_3)\ (y_2)\ (y_4)$	3	2	1	0	0
16	y_3	y_2	y_4	y_1	$(y_1, y_3, y_4)\ (y_2)$	2	1	0	1	0
17	y_3	y_4	y_1	y_2	$(y_1, y_3)\ (y_2, y_4)$	2	0	2	0	0
18	y_3	y_4	y_2	y_1	(y_1, y_3, y_2, y_4)	1	0	0	0	1
19	y_4	y_1	y_2	y_3	(y_1, y_4, y_3, y_2)	1	0	0	0	1
20	y_4	y_1	y_3	y_2	$(y_1, y_4, y_2)\ (y_3)$	2	1	0	1	0
21	y_4	y_2	y_1	y_3	$(y_1, y_4, y_3)\ (y_2)$	2	1	0	1	0
22	y_4	y_2	y_3	y_1	$(y_1, y_4)\ (y_2)\ (y_3)$	3	2	1	0	0
23	y_4	y_3	y_1	y_2	(y_1, y_4, y_2, y_3)	1	0	0	0	1
24	y_4	y_3	y_2	y_1	$(y_1, y_4)\ (y_2, y_3)$	2	0	2	0	0

Table 2 – Distributions of quantities of values $\xi_n = k$ and $\chi_{n,L} = k$ for all substitutions from S_4

k	0	1	2	3	4
$\#(\xi_n = k)$	0	6	11	6	1
$\#(\chi_{n,L=1} = k)$	9	8	6	0	1
$\#(\chi_{n,L=2} = k)$	15	6	3	0	0
$\#(\chi_{n,L=3} = k)$	16	8	0	0	0
$\#(\chi_{n,L=4} = k)$	18	6	0	0	0

Estimation of the Probability of the Cycle (y_i)

Consider now the substitutions containing k = 2 cycles of length $L=1$. We have 6 substitutions (table 4), of which three substitutions in a cyclic decomposition contain cycles $(y_1)\ (y_{i\neq 1})$, three substitutions contain cycles $(y_2)\ (y_{i\neq 2})$, three substitutions contain cycles $(y_3)\ (y_{i\neq 3})$ and three sub-

stitutions contain cycles (y_4) $(y_{i\neq 4})$. It is clear that one and the same substitution can be assumed to different ways, i.e. it can, in its cyclic decomposition, be treated to substitutions containing cycles (y_i) $(y_{j\neq i})$ and substitutions containing cycles (y_j) $(y_{i\neq j})$. For example, the sixth substitution has a cyclic decomposition (y_1) (y_2, y_4) (y_3), it should be considered as a substitution with cycles (y_1) $(y_{j\neq 1})$ and with substitutions with cycles (y_3) $(y_{i\neq 3})$.

Table 3 – Substitutions containing 1 cycle of length 1 and one cycle of length 3

№	Result of substitution				Decomposition of substitution to cycles
	$s(y_1)$	$s(y_2)$	$s(y_3)$	$s(y_4)$	
4	y_1	y_3	y_4	y_2	(y_1) (y_2, y_3, y_4)
5	y_1	y_4	y_2	y_3	(y_1) (y_2, y_4, y_3)
9	y_2	y_3	y_1	y_4	(y_1, y_2, y_3) (y_4)
12	y_2	y_4	y_3	y_1	(y_1, y_2, y_4) (y_3)
13	y_3	y_1	y_2	y_4	(y_1, y_3, y_2) (y_4)
16	y_3	y_2	y_4	y_1	(y_1, y_3, y_4) (y_2)
20	y_4	y_1	y_3	y_2	(y_1, y_4, y_2) (y_3)
21	y_4	y_2	y_1	y_3	(y_1, y_4, y_3) (y_2)

Table 4 – Substitutions containing 2 cycle of length 1 and one cycle of length 2

№	Result of substitution				Decomposition of substitution to cycles
	$s(y_1)$	$s(y_2)$	$s(y_3)$	$s(y_4)$	
2	y_1	y_2	y_4	y_3	(y_1) (y_2) (y_3, y_4)
3	y_1	y_3	y_2	y_4	(y_1) (y_2, y_3) (y_4)
6	y_1	y_4	y_3	y_2	(y_1) (y_2, y_4) (y_3)
7	y_2	y_1	y_3	y_4	(y_1, y_2) (y_3) (y_4)
15	y_3	y_2	y_1	y_4	(y_1, y_3) (y_2) (y_4)
22	y_4	y_2	y_3	y_1	(y_1, y_4) (y_2) (y_3)

The number of substitutions containing k = 2 cycles (y_i) $(y_{j\neq i})$ of length $L=1$ for fixed $y_i \in Y$ are calculated as follows. In total, there are exactly $C_{n=4}^{kL=2}=6$ ways to simultaneously select values $y_i, y_{j\neq i} \in Y$. But for fixed $y_i \in Y$ there are exactly $C_{n-1=3}^{kL-1=1}=3$ ways for choosing a value $y_{j\neq i} \in Y$. That is, the total number of substitutions containing k = 2 cycles of length $L=1$ must be multiplied by the value $\frac{C_{n-1=3}^{kL-1=1}}{C_{n=4}^{kL=2}}=\frac{3}{6}$, we obtain the desired value, for an arbitrary fixed $y_i \in Y$ the number of substitutions containing k = 2 cycles (y_i) $(y_{j\neq i})$ of length $L=1$, is equal to three. For example, for fixed value $y_1 \in Y$ the second, third, and sixth substitutions contain k = 2 cycles of length $L=1$ with cyclic decompositions: (y_1) (y_2) (y_3, y_4), (y_1) (y_2, y_3) (y_4) та (y_1) (y_2, y_4) (y_3).

Consider substitutions containing k = 4 cycles of length $L=1$ (no substitution can have k = 3 cycles of length $L=1$). We have 1 substitution (Table 1), its cyclic decomposition has the form: (y_1) (y_2) (y_3) (y_4). Applying the same formula, we have $\frac{C_{n-1=3}^{kL-1=3}}{C_{n=4}^{kL=4}}=\frac{1}{1}=1$, that is, the total number of

substitutions containing four cycles of length 1 coincides with the number of substitutions with cyclic decomposition (y_i) $(y_{j\neq i})$ $(y_{u\neq i,j})$ $(y_{v\neq i,j,u})$, as it should be.

We will calculate the number of substitutions from S_4 (Table 1), which for fixed $y_i \in Y$ necessarily contain a cycle (y_i) of length (y_i). To do this, we must summarize the number of substitutions that for fixed $y_i \in Y$ in their cyclic decomposition contain a different number of cycles of length $L=1$, namely, 2 substitutions with k = 1 cycle (y_i), three substitutions with k = 2 cycles (y_i) $(y_{j\neq i})$ and one substitution with k = 4 cycles (y_i) $(y_{j\neq i})$ $(y_{u\neq i,j})$ $(y_{v\neq i,j,u})$. In general, we have 6 substitutions from the total of 24 substitutions of the symmetric group. That is, at randomly equal probable selection of substitutions from S_4 the probability that in it for arbitrary fixed $y_i \in Y$ will be observed cycle (y_i) of length $L=1$ is equal to 6/24 = 1/4.

Estimation of probability of occurrence of a cycle $(y_i, y_{j\neq i})$

Consider the case when for arbitrary fixed $y_i \in Y$ randomly chosen substitution s from S_4 will necessarily contain a cycle $(y_i, y_{j\neq i})$ of length $L=2$.

First we consider substitutions containing k = 1 cycle of length $L=2$. We have 6 such substitutions, which are given in Table 4 (if the substitution from S_4 contains two cycles of length 1, then it necessarily contains one cycle of length 2). We will calculate the number of substitutions, which for an arbitrary fixed $y_i \in Y$ contain k = 1 cycle of length $L=2$ of form $(y_i, y_{j\neq i})$. In total, there are exactly $C_{n=4}^{kL=2}=6$ ways to simultaneously select values $y_i, y_{j\neq i} \in Y$. But for each $y_i \in Y$ there is exactly $C_{n-1=3}^{kL-1=1}=3$ ways to choose a value $y_{j\neq i} \in Y$. That is, the number of substitutions, which for an arbitrary fixed $y_i \in Y$ in a cyclic decomposition contain k = 2 cycles (y_i) $(y_{j\neq i})$ of length l=1, is equal to $6 \cdot \frac{C_{n-1=3}^{kL-1=1}}{C_{n=4}^{kL=2}}=3$. For example, for $y_1 \in Y$ the seventh, fifteenth and twenty second substitutions contain k=1 cycles of length l=2 each.

Let's also consider substitutions containing k = 2 cycles of length $L=2$. In total there are 3 such substitutions in S_n, this (see Table 1):

- the eighth substitution with a cyclic decomposition (y_1, y_2) (y_3, y_4);
- seventeenth substitution with cyclic decomposition (y_1, y_3) (y_2, y_4);
- last, twenty fourth substitution with cyclic decomposition (y_1, y_4) (y_2, y_3).

The number of substitutions, which for any fixed $y_i \in Y$ contains k = 2 cycles $(y_i, y_{j\neq i})$ and $(y_{u\neq i,j}, y_{v\neq i,j,u})$ of length $L=2$, is calculated in the same way. In total, there are exactly $C_{n=4}^{kL=4}=1$ ways to simultaneously select values $y_i, y_{j\neq i}, y_{u\neq i,j}, y_{v\neq i,j,u} \in Y$. This choice is no longer limited, because for selected $y_i, y_{j\neq i}, y_{u\neq i,j}, y_{v\neq i,j,u} \in Y$ corresponding k=2 cycles $(y_i, y_{j\neq i})$ та $(y_{u\neq i,j}, y_{v\neq i,j,u})$ are defined unambiguously (according to the formula there are $C_{n-1=3}^{kL-1=3}=1$ variants). That is, the number of substitutions, which for an arbitrary fixed value $y_i \in Y$ have cyclic decomposition $(y_i, y_{j\neq i})$ $(y_{u\neq i,j}, y_{v\neq i,j,u})$ coincides with the total number of substitutions with k = 2 cycles of length $L=2$, i.e. equal to 3. No substitution from S_4 can have k = 3 and k = 4 cycles of length $L=2$. We immediately proceed to calculate the probability of occurrence of the cycle $(y_i, y_{j\neq i})$ in randomly selected substitutions. We sum up the number of substitutions, which for an arbitrary fixed $y_i \in Y$ in their cyclic decomposition contain a different number of cycles of length

$L = 2$, namely, we have 3 substitutions with $k = 1$ cycle $(y_i, y_{j\neq i})$, and three substitutions with $k = 2$ cycles $(y_i, y_{j\neq i})(y_{u\neq i,j}, y_{v\neq i,j,u})$. In general, we have 6 substitutions from the total of 24 substitutions of the symmetric group, that is, at randomly equal probable selection of substitutions from S_4 the probability that in it for arbitrary fixed $y_i \in Y$ will be observed cycle $(y_i, y_{j\neq i})$ of length $L = 2$ is equal to 6/24=1/4.

Estimation of probability of occurrence of a cycle $(y_i, y_{j\neq i}, y_{u\neq i,j})$

Similar to the above, we will calculate the number of substitutions, which for arbitrary fixed $y_i \in Y$ have cycle $(y_i, y_{j\neq i}, y_{u\neq i,j})$. In an arbitrary substitution $s \in S_4$, there can be no more than one such cycle, i.e. cases with k>1 are impossible. Each cycle of length $L = 3$ in the substitution decomposition is combined with a cycle of length 1, i.e. all eight of these substitutions are given in Table 3. We will calculate the number of substitutions which for arbitrary fixed $y_i \in Y$ necessarily contain a cycle $(y_i, y_{j\neq i}, y_{u\neq i,j})$. In total, there are exactly $C_{n=4}^{kL=3} = 4$ ways to simultaneously select values $y_i, y_{j\neq i}, y_{u\neq i,j} \in Y$. But for each $y_i \in Y$ there are exactly $C_{n-1=3}^{kL-1=2} = 3$ ways to choose of values $y_{j\neq i}, y_{u\neq i,j} \in Y$. That is, the number of substitutions, which for an arbitrary fixed value $y_i \in Y$ necessarily have a cycle $(y_i, y_{j\neq i}, y_{u\neq i,j})$, is defined as $8 \cdot \frac{C_{n-1=3}^{kL-1=2}}{C_{n=4}^{kL=3}} = 6$. For example, for $y_1 \in Y$ these are 6 substitutions (№ 9, 12, 13, 16, 20, 21) with cyclic decompositions (see Table 3): $(y_1, y_2, y_3)(y_4)$, $(y_1, y_2, y_4)(y_3)$, $(y_1, y_3, y_2)(y_4)$, $(y_1, y_3, y_4)(y_2)$, $(y_1, y_4, y_2)(y_3)$ and $(y_1, y_4, y_3)(y_2)$.

Consequently, at randomly equal probable selection of substitutions from S_4 the probability that in it for arbitrary fixed $y_i \in Y$ will be observed cycle $(y_i, y_{j\neq i}, y_{u\neq i,j})$ of length $L = 3$ is equal to 6/24 = 1/4.

Estimation of Probability of Occurrence of a Cycle $(y_i, y_{j\neq i}, y_{u\neq i,j}, y_{v\neq i,j,u})$

In any substitution $s \in S_4$ there can be only one cycle of length $L = 4$. Such substitutions are given in Table 5.

Table 5 – Substitutions containing 1 cycle of length 4

№	Result of substitution				Decomposition of substitution to cycles
	$s(y_1)$	$s(y_2)$	$s(y_3)$	$s(y_4)$	
10	y_2	y_3	y_4	y_1	(y_1, y_2, y_3, y_4)
11	y_2	y_4	y_1	y_3	(y_1, y_2, y_4, y_3)
14	y_3	y_1	y_4	y_2	(y_1, y_3, y_4, y_2)
18	y_3	y_4	y_2	y_1	(y_1, y_3, y_2, y_4)
19	y_4	y_1	y_2	y_3	(y_1, y_4, y_3, y_2)
23	y_4	y_3	y_1	y_2	(y_1, y_4, y_2, y_3)

Obviously, all such substitutions necessarily have in their single cycle all elements from set $Y = \{y_1, y_2, y_3, y_4\}$, i.e. for each $y_i \in Y$ in each substitution of Table 5 there will be a cycle $(y_i, y_{j\neq i}, y_{u\neq i,j}, y_{v\neq i,j,u})$. Check the above formula: $\frac{C_{n-1=3}^{kL-1=3}}{C_{n=4}^{kL=4}} = 1$, indeed, all substitutions from Table 5 will necessarily contain a cycle $(y_i, y_{j\neq i}, y_{u\neq i,j}, y_{v\neq i,j,u})$ for each arbitrary fixed $y_i \in Y$. Consequently, the probability of a cycle $(y_i, y_{j\neq i}, y_{u\neq i,j}, y_{v\neq i,j,u})$ in a randomly selected substitution $s \in S_4$ is equal to 6/24 = 1/4.

6. Interpretation of the received results to the properties of BSC

The obtained analytic expressions (15) and (16) allow us to estimate the number of substitutions from the symmetric group, which for a certain element of the set of transformations necessarily contain a cycle of a certain length with this element, and the corresponding probability to randomly select substitution with such cycle.

We use these formulas to study periodic properties of OFB mode. In this case, we assume that the probabilistic properties of substitutions generated by the encryption function correspond to certain properties of random substitution, i.e. they correspond to our representations of such "ideal" BSC, which at any given encryption key randomly and equally compares any encrypt text to any open text. In this way, the l-bit BSC will implement a subset of the symmetric group of substitutions of degree 2^l, selecting a particular substitution s from S_{2^l} associated with the entered encryption key. On the all set of cipher keys we can choose substitution, which for an arbitrary fixed $y_i = S \in Y = \{y_1, y_2, ..., y_{2^l}\}$ necessarily has a cycle $(y_i, s_i(y_i), s_i^2(y_i), ..., s_i^{L-1}(y_i))$ of length $L = l_i \in \{1, 2, ..., n\}$. The probability of this event is determined by (16).

In practice this means that the probability of a cycle of any length in randomly chosen substitution from a symmetric group S_{2^l} for an arbitrary fixed element of a set does not depend on either this element or the length of the cycle. It depends only on the order $n = 2^l$ of the substitutions of the symmetric group S_n and is defined as the inverse value, that is equal to 1 / n. For any fixed value of the introduced initialization vector $y = S$, the probability that the corresponding output block $\gamma_i, i = 0,1,..,n-1$ formed in the OFB mode will have a period of length $L = l_i \in \{1, 2, ..., n\}$ does not depend on either the value of this initialization vector or the length of the period. It is determined only by the degree of substitution, that is, in this case, the digit of the cipher and it is equal to 2^{-l}.

Determine the probability that the period of formed output blocks will be not less than 2^m blocks, i.e. the probability of such an event, when for fixed value of the initialization vector $y_i = S$ the corresponding output blocks will not be repeated during 2^m iterations by the formulas (1) and (3) when forming the output blocks γ_i:

$$P(\forall i, j \in \{1, 2, ..., 2^m\} : \gamma_i \neq \gamma_j \big|_{i \neq j}) = 1 - \sum_{L=1}^{2^m} \sum_{k=1}^{[2^l/L]} P(\chi_{2^l, L} = k) \frac{C_{2^l-1}^{kL-1}}{C_{2^l}^{kL}} = \sum_{L=2^m+1}^{2^l} \sum_{k=1}^{[2^l/L]} P(\chi_{2^l, L} = k) \frac{C_{2^l-1}^{kL-1}}{C_{2^l}^{kL}}. \quad (17)$$

Taking into account (16) we obtain:

$$P(\forall i, j \in \{1, 2, ..., 2^m\} : \gamma_i \neq \gamma_j \big|_{i \neq j}) = 1 - \sum_{L=1}^{2^m} 2^{-l} = \sum_{L=2^m+1}^{2^l} 2^{-l} = 1 - 2^{m-l}. \quad (18)$$

Thus, when the assumption about the correspondence of certain probabilistic properties of the cipher to the properties of a random substitution is correct, the probability of non-repetition of output block at a certain length is a function of this length. This fact determines the main limitation of the use of the OFB mode, it is directly derived from the results of the research. The main limitation on the use of the OFB mode for BSC "Kalyna" [10,11] is specified in Appendix G.2 "Limit on the total length of messages protected by the use of one key" namely:

- when the block size is equal to 128 bits, it is recommended to limit the number of blocks protected by the single key to value 2^{60} (16 million TB);
- when the block size is equal to 256 bits, it is recommended to limit the number of blocks protected by the single key to value 2^{124};
- when the block size is equal to 512 bits, it is recommended to limit the number of blocks protected by the single key to value 2^{251}.

Since, as shown in Section 2, the essence of protecting an informational message according to

the OFB mode specification is in addition of output blocks to it, then the restrictions specified in appendix G.2 relate to the restrictions on the length of output blocks, which are formed by multiple encryption of the same non-secret initialization vector by the formulas (1), (3). That is, the implementation of the restrictions recommended by the specification of the national standard provides certain probabilistic indicators of non-periodic output blocks, namely:

- when the block size is l = 128 bits, and when performing the recommended limitation of the number of blocks protected by a single key with the size $2^m = 2^{60}$, the probability of non-repetition of the output blocks $1-2^{m-l} = 1-2^{-68} > 1-2^{-64}$ will be ensured;
- when the block size is l = 256 bits, and when performing the recommended limitation of the number of blocks protected by a single key with the size $2^m = 2^{124}$, the probability of non-repetition of the output blocks $1-2^{m-l} = 1-2^{-132} > 1-2^{-128}$ will be ensured;
- when the block size is l = 512 bits, and when performing the recommended limitation of the number of blocks protected by a single key with the size $2^m = 2^{251}$, the probability of non-repetition of the output blocks $1-2^{m-l} = 1-2^{-261} > 1-2^{-256}$ will be ensured.

More often, in the theory of information security, the inverse value is used, it is the probability that a formed output blocks with a length that does not exceed a certain limit will have at least one repetition. Taking into account (17) and (18), this probability will be determined as:

$$P_{l,m} = 1 - P(\forall i, j \in \{1,2,...,2^m\} : \gamma_i \neq \gamma_j \big|_{i \neq j}) = \sum_{L=1}^{2^m} \sum_{k=1}^{[2^l/L]} P(\chi_{2^l,L} = k) \frac{C_{2^l-1}^{kL-1}}{C_{2^l}^{kL}} = 2^{m-l}. \qquad (19)$$

Figure 2 shows the dependence $P_{l,m}$ and m for different l ($\gamma_i, i = 0,1,..,n-1$ *with the length not more than* $2^m = 2^{60}$ *blocks*).

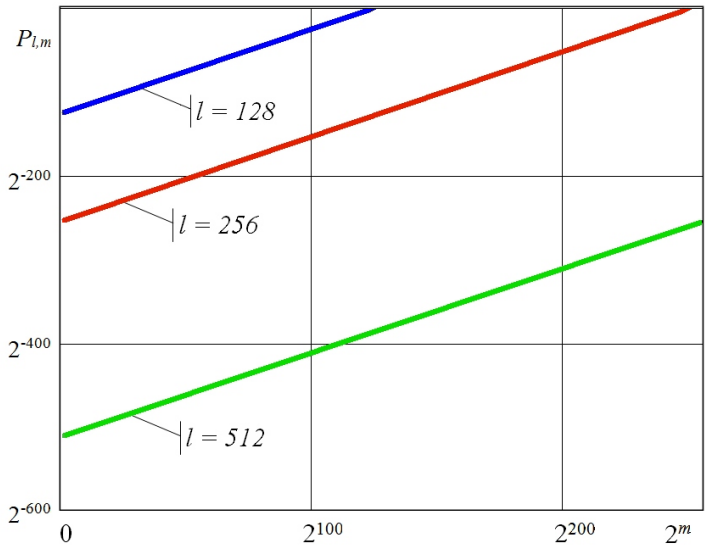

Fig. 2 – The dependence of the probability that there will be at least one repetition in output blocks

From the dependencies shown in Fig. 2, it is evident that increasing the length of the output blocks leads to an increase in the probability of any number of repetitions of the output blocks. These graphs can be used to justify certain restrictions, for example, if you want to reduce the probability of any number of repetitions of the output blocks, you must reduce its length.

7. Conclusions

Based on the obtained results, we can do conclusions that are important in practical terms.

1. Properties of modern symmetric cryptotransformations depend not only on the characteristics of BSC, but also on the mode of application. Therefore the National Standard "Information technology. Cryptographic protection of information. The algorithm of symmetric block transformation" provides 10 modes of cryptotransformations: simple substitution (Electronic Codebook - basic transformation), Counter, Cipher Feedback, Symmetric Key Block Cipher-Based Message Authentication Code, Cipher Block Chaining, Output Feedback, Galois/Counter Mode and Galois Message Authentication Code, Counter with Cipher Block Chaining-Message Authentication Code, XOR Encrypt XOR (XEX) Tweakable Block Cipher, Key Wrapping.

2. In OFB mode, which is used to provide a privacy service, the output message is protected by

addition of output blocks, which are formed by multiple encryption of the same non-secret initialization vector. If we assume that the properties of the cipher correspond to certain properties of random substitution, then the periodicity of the output blocks will be determined by the presence of cycles in randomly selected substitutions from the symmetric group, and the selection of the substitution is set to the value of the secret key.

3. Investigation of the properties of randomly selected substitution s from the symmetric group S_n has shown that the probability of cycle $s_i = (y_i, s_i(y_i), s_i^2(y_i), ..., s_i^{l_i-1}(y_i))$ of any length $L = l_i$ for an arbitrary fixed element y_i from the set $Y = \{y_1, y_2, ..., y_n\}$ does not depend on either this element y_i or the length $L = l_i$ of the cycle. This probability depends only on the order *n* of the substitutions of the symmetric group S_n and is defined as an inverse value, that is equal to 1 / n.

4. Thus, the periodic properties of the output blocks in OFB mode are determined by the distribution of the probabilities of the number of cycles of random substitution. The selection of a secret key that parameterizes the encryption function corresponds to the selection of a particular substitution from the symmetric group; selecting the initialization vector value corresponds to the selection of an element y_i from the set of elements $Y = \{y_1, y_2, ..., y_n\}$ over which substitution occurs. But neither the actual value of the initialization vector nor the length of the output blocks period has any effect on the probability of obtaining an output blocks of a certain period. This probability is determined only by the degree $n = 2^l$ of substitution, that is, by the size l of the basic cipher transformation.

5. From the point of view of the practical application of symmetric cryptographic transformations to output blocks, the requirements of homogeneity are proposed at a length not exceeding the established limit. The probability of such event is determined by $1 - 2^{m-l}$ where 2^m is the length restriction of the output blocks. For example, for a 128-bit cipher "Kalyna", when we have limiting the length of a output blocks to 2^m in the OFB mode, the probability that the gamma blocks never coincide equals to $1 - 2^{m-l} = 1 - 2^{-68}$, i.e. it is much more than $1 - 2^{-64}$. The probability that at length no more 2^m blocks of output blocks in the OFB mode will coincide at least once, will equal $P_{l,m} = 2^{m-l}$. This dependence can be used to substantiate the restrictions on the length of the output blocks when the upper limit of probability $P_{l,m}$ is set.

6. The specification of BSC "Kalyna" [10,11] recommended certain restrictions on the total length of messages protected by the use of one key. As for OFB mode such restrictions should be considered as requirements for the maximum length of output blocks, which with a certain probability will not be repeated. The above recommendations are fundamental because the occurrence of output blocks repetition is the most dangerous case when using the OFB mode, since in this case, the attacker will almost certainly violate the established privacy mode of the messages. For example, if the probability of a repetition of the output blocks is equal to 2^{-64}, then the length of output blocks for 128-bits BSC "Kalyna" should not exceed 2^{64} blocks (in the standard this restriction is more stringent and equal to 2^{60}).

7. The estimations of the probability of forming the output blocks with given period can be considered as a criterion for selection of cryptographic primitives, or criterion of the statistical test. Indeed, if the studied cryptographic primitive in OFB mode with equal probability forms output blocks with any period and this probability is determined by inverse to the degree of substitution, then by the cyclical properties this cryptographic primitive responsible to probabilistic properties of random substitution and by this criterion can be adopted for use. Actually such studies, particularly on nonlinear replacement nodes or reduced BSC models are promising direction for further work.

References

[1] Sachkov V.N. Introduction to combinatorial methods of discrete mathematics/ V.N. Sachkov. – Moskow: Nauka; Ed. Ph.-math. Lit., 1982. – 384 p.
[2] Tronin S.N. Introduction to the theory of groups/ S.N. Tronin. – Kazan: Kazan State University, 2006. – 100 p.
[3] Alexandrov P.S. Introduction to the theory of groups/ P.S. Alexandrov. – Moskow: Nauka, 1980. – 145 p.

[4] Menezes Alfred J. Handbook of Applied Cryptography/ Alfred J. Menezes, Paul C. van Oorschot, Scott A. Vanstone. – CRC Press, 1997. – 794 p.
[5] Dolgov V.I. Analysis of cyclic properties of block ciphers/ V.I. Dolgov, I.V. Lisitskaya, V.I. Ruzhentsev// Applied radio electronics . – 2007. – Vol.6. – №2. – P. 257 – 263.
[6] Kuznetsov A.A. Linear properties of block symmetric ciphers presented to the Ukrainian competition/ A.A. Kuznetsov, I.V. Lisitskaya, S.A. Isaev //Applied radio electronics. – 2001. – Vol. 10. – № 2. – P.135 – 140.
[7] Soroka L.S. Investigation of the differential properties of block-symmetric ciphers/ L.S. Soroka, A.A. Kuznetsov, I.V. Moskovchenko, S.A. Isaev // Information Processing Systems. – 2010. – Vol. 6 (87). – P. 286 – 294.
[8] Dolgov V.I. Block symmetric ciphers are random substitutions. Combinatorial indices/ V.I. Dolgov, M.Yu. Rodinko //Applied radio electronics. – 2013. – Vol.12. – № 2. – P. 236 – 239.
[9] Kuznetsov O.O. Analysis of collision properties of Galois Message Authentication Code with selective Counter/ O.O. Kuznetsov, D.V. Ivanenko, Ie.P. Kolovanova // Bulletin of V. Karazin Kharkiv National University. Series "Mathematical Modelling. Information Technology. Automated Control Systems". – 2014. – № 1097. – Issue 23. – P. 55 – 71.
[10] Information Technology. Cryptographic protection of information. Symmetric block algorithm transformation: DSTU. – Kyiv: Ministry of Economic Development of Ukraine, 2015. – 238 p.
[11] Gorbenko I.D. Development of a new symmetric block cipher: Report on the first phase of research "Algorithm" (intermediate) / I.D. Gorbenko :in 4 t. – T. 4. – Kharkiv: JSC «IIT», 2014. – 304p.

Рецензент: Сергей Толюпа, д.т.н., проф., Киевский национальный университет имени Т. Шевченко, Киев, Украина.
E-mail: tolupa@i.ua

Поступила: Май 2017.

Авторы:
Александр Кузнецов, д.т.н., проф., академик Академии наук прикладной радиоэлектроники, ХНУ имени В.Н. Каразина, Харьков, Украина.
E-mail: kuznetsov@karazin.ua
Евгения Колованова, к.т.н., ст. преподаватель, ХНУ имени В.Н. Каразина, Харьков, Украина.
E-mail: e.kolovanova@gmail.com
Татьяна Кузнецова, научный сотрудник каф. безопасности информационных систем и технологий (БИСТ), ХНУ имени В.Н. Каразина, Харьков, Украина.
E-mail: kuznetsova.tatiana17@gmail.com

Периодические свойства шифргаммы в режиме Output Feedback.

Аннотация: Исследуются свойства режима гаммирования с обратной связью по шифргамме (анг. – Output Feedback). С применением математического аппарата теории подстановок исследуются периодические свойства гаммы, в частности, проводится оценка вероятности появления гаммы определенного периода при условии соответствия свойств шифра определенным свойствам случайной подстановки. Разрабатываются практические рекомендации по применению режима гаммирования с обратной связью по шифргамме, обосновываются требования и ограничения, вытекающие из полученных оценок периодических свойств гаммы.

Ключевые слова: режим шифрования, периодичность гаммы, случайная подстановка, Output Feedback.

Рецензент: Сергій Толюпа, д.т.н., проф., Київський національний університет імені Т. Шевченка, Київ, Україна.
E-mail: tolupa@i.ua

Надійшло: Травень 2017.

Автори:
Олександр Кузнецов, д.т.н., проф., академік Академії наук прикладної радіоелектроніки, ХНУ імені В.Н. Каразіна, Харків, Україна.
E-mail: kuznetsov@karazin.ua
Євгенія Колованова, к.т.н., ст. викладач, ХНУ імені В.Н. Каразіна, Харків, Україна.
E-mail: e.kolovanova@gmail.com
Тетяна Кузнецова, науковий співробітник каф. безпеки інформаційних систем і технологій (БІСТ), ХНУ імені В.Н. Каразіна, Харків, Україна.
E-mail: kuznetsova.tatiana17@gmail.com

Періодичні властивості шифргами у режимі Output Feedback.

Анотація: Досліджуються властивості режиму гамування зі зворотнім зв'язком за шифргамою (анг. – Output Feedback). Із застосуванням математичного апарату теорії підстановок досліджуються періодичні властивості гами, зокрема проводиться оцінка ймовірності появи гами певного періоду за умови відповідності властивостей блокового симетричного шифру певним властивостям випадкової підстановки. Розробляються практичні рекомендації щодо застосування режиму гамування зі зворотнім зв'язком за шифргамою, обґрунтовуються вимоги та обмеження, що випливають із отриманих оцінок періодичних властивостей гами.

Ключові слова: режим шифрування, періодичність гами, випадкова підстановка, Output Feedback.