The branch numbers of linear transformations in encryption algorithms

A.V. Erokhin, F.M. Malyshev, A.E. Trishin

C. Shannon formulated essential confusion and diffusion requirements for encryption algorithms. By the 60th years of the twentieth century his recommendations were implemented in the SP-ciphers.

The linear cryptanalysis and its dual differential cryptanalysis led to appearance of new block ciphers including XSL-ciphers. The diffusion property may be measured by a numerical characteristic called *the linear medium's coefficient of diffusion* (LMCD) of cipher transformation.

Actually the linear medium of a cipher has two coefficients of diffusion associated with the linear and differential cryptanalysis. The related terms are *the linear medium's linear coefficient of diffusion* (LMLCD) and *the linear medium's differential coefficient of diffusion* (LMDCD) of cipher transformation. In this presentation we will consider only the LMLCD (in short, LMCD).

Linear transformations (and their matrices) in ciphers are characterized by the branch numbers.

Unfortunately the linear and differential branch numbers do not distinguish substitution matrices used in SP-ciphers. All permutations have a minimal value of the branch number equal to 2 but different permutations have different diffusion properties (according to Shannon) as is demonstrated by the avalanche effect.

We want to eliminate this disadvantage by using the multidimensional linear cryptanalysis.

1.1. The functional scheme defining the cipher transformation

$$F : V_N \times V_K \to V_M,\ F(a, z) = b,\ (a, z) \in V_N \times V_K,$$

may be represented by the command sequence of its program realization.

Here $a \in V_N$ is a block of plaintext, $b \in V_M$ is a block of ciphertext, $z \in V_K$ is a secret key. In the case of block ciphers $M = N$.

Let

$$f_i : V_{n_i} \to V_{m_i},\ f_i(x_i) = y_i,\ i = 1, \ldots, k,$$

be nonlinear functional elements of the functional scheme defining the cipher transformation. All other linear operations of the functional scheme form the linear medium of cipher transformation.

We may assume that the mappings f_i are arranged so that outputs of f_i may be inputs of f_j (perhaps indirectly, after linear operations) only if $i < j$. As a result,

$$x_j = c_j(z, a, y_1, \ldots, y_{j-1}) = zc_{*j} + ac_{0j} + y_1c_{1j} + \ldots + y_{j-1}c_{j-1,j},$$

$$b = c_{k+1}(z, a, y_1, \ldots, y_k) = zc_{*,k+1} + ac_{0,k+1} + \ldots + y_kc_{k,k+1},$$

where $c_j : V_{K+N+\sum_{i=1}^{j-1} m_i} \to V_{n_j},\ j = 1, \ldots, k, k+1$, are linear mappings.

Here $c_{ij},\ i = 0, 1, \ldots, j-1,\ j = 1, \ldots, k, k+1$, are a $m_i \times n_j$ matrices, $m_0 = N$, $n_{k+1} = M$. Further, $c_{*j}, j = 1, \ldots, k, k+1$, are a $K \times n_j$ matrices.

The linear mappings c_j are combined into united linear mapping

$$C : V_{K+N+\sum_{i=1}^{k} m_i} \to V_{\sum_{i=1}^{k} n_i+M}$$

and C is represented by $(K + N + \sum_{i=1}^{k} m_i) \times (\sum_{i=1}^{k} n_i + M)$ matrix.

The uppermost "row" of matrix C is the matrix $C_0 = = (c_{*1}, \dots, c_{*k}, c_{*,k+1})$. Let $\widetilde{C}$ be a submatrix of matrix C consisting of "rows" $(c_{i1}, \dots, c_{ik}, c_{i,k+1})$, $i = 0, 1, \dots, k$ (suppose $c_{ij} = 0$ for $i \geqslant j$). Then we have equation

$$(z, a, y_1, \dots, y_k)\, C = (x_1, \dots, x_k, b),$$

or

$$(a, y)\widetilde{C} + zC_0 = (z, a, y)\, C = (x, b),$$

where $(y_1, \dots, y_k) = y$, $(x_1, \dots, x_k) = x$.

1.2. Additive method of constructing s-dimensional linear relations.

Suppose $s \geqslant 1$. A multidimensional linear relation of cipher transformation is given by linear mappings $L' : V_N \to V_s$, $L : V_K \to V_s$, $L'' : V_M \to V_s$ and is represented as

$$\eta = aL' + zL + bL'',$$

where $b = F(a, z)$, a is a uniformly distributed random vector, $z \in V_K$ is a fixed key.

An efficiency of the relation is characterized by probability distribution of the vector η on the set V_s. We may use the entropy $H(\eta)$ as a measure of uncertainty of the vector η.

The relation η is obtained by summing the local s-dimensional probability linear relations of mappings f_i for all $i = 1, \ldots, k$, namely

$$\eta_i = x_i l'_i + y_i l''_i,$$

where $y_i = f_i(x_i)$. A probability distribution of η_i is calculated under the assumption that $x_i \in V_{n_i}$ are uniformly distributed.

The entire set of relations η_i is given by the set $\mathfrak{L} = ((l'_i, l''_i), i = 1, \ldots, k)$ which consists of binary $n_i \times s$ and $m_i \times s$ matrices defining linear mappings $l'_i : V_{n_i} \to V_s$, $l''_i : V_{m_i} \to V_s$, $i = 1, \ldots, k$.

We will call this set $\mathfrak{L}$ as the system of the local s-dimensional probability linear relations of cipher transformation F.

The mappings l'_i, l''_i, $i = 1, ..., k$, must satisfy two requirements.

The first requirement is to move the distribution of vectors $\eta_i = = x_i l'_i + y_i l''_i$ near to degenerate distribution, that is to make these vectors more specific, therefore, in particular $\mathrm{Im} l'_i \subseteq \mathrm{Im} l''_i$.

The second requirement is the conformity of the system $\mathfrak{L} = = ((l'_i, l''_i), i = 1, \ldots, k)$, so we can reduce (but without using equations $y_i = f_i(x_i)$, $i = 1, \ldots, k$) the sum

$$\eta_{\mathfrak{L}} = \sum_{i=1}^{k} \eta_i = \sum_{i=1}^{k} (x_i l'_i + y_i l''_i)$$

to the form

$$\eta = aL' + zL + bL'',$$

where $L' : V_N \to V_s$, $L : V_K \to V_s$, $L'' : V_M \to V_s$ are some linear mappings.

The second requirement is equivalent to the solvability of the equation

$$\widetilde{C}\begin{pmatrix} l' \\ L'' \end{pmatrix} = \begin{pmatrix} L' \\ l'' \end{pmatrix}$$

with respect to $N \times s$ and $M \times s$ matrices L', L''. In this equation the matrices l', l'' consist of stacked matrices l'_i, l''_i, $i = 1, \dots, k$, respectively.

If the system $\mathfrak{L}$ is conformal then we suppose $L = C_0 \begin{pmatrix} l' \\ L'' \end{pmatrix}$, and

$$\eta_{\mathfrak{L}} = xl' + yl'' = \sum_{i=1}^{k}(x_i l'_i + y_i l''_i) = \sum_{i=1}^{k} \eta_i = aL' + zL + bL'' = \eta.$$

The set $\mathfrak{W}_s$ of all conformal systems $\mathfrak{L} = (l', l'')$ is a vector space over the field $GF(2)$.

The efficiency of a key recovery depends on a value

$$\sigma = \sum_{v \in V_s} \varepsilon_v^2 \approx \ln 2 \cdot (s - H(\eta))/2^{s-1},$$

where $\left\{p_v = \frac{1}{2^s} + \varepsilon_v,\, v \in V_s\right\}$ is a probability distribution of η. The less the uncertainty of $H(\eta)$ (or the greater σ), the less amount of data is needed for a key recovery attack.

This probability distribution is estimated under the assumption that random summands η_i, $i = 1, \ldots, k$, are statistically independent and x_i, $i = 1, \ldots, k$, are uniformly distributed.

Therefore we are interested in numbers $i \in \{1, \ldots, k\}$ such that $x_i l_i' = f_i(x_i) l_i''$ for all $x_i \in V_{n_i}$, particularly $l_i' = 0, l_i'' = 0$. For such numbers a random vector η_i does not introduce an uncertainty into η. Thus conformal systems $\mathfrak{L}$ having a minimal value of

$$\theta_{\mathfrak{L}} = \left|\{i \in \{1, \ldots, k\} \,|l_i'' \neq 0\}\right|$$

are preferred.

2.1. The linear medium's coefficient of diffusion.

In the case $s = 1$ the LMCD of cipher transformation with linear medium C is determined by the formula

$$\theta_1(C) = \min_{\mathfrak{L} \in \mathfrak{W}_1 \setminus \{0\}} \theta_{\mathfrak{L}}.$$

Let the matrix $\Lambda \in GL(n, 2)$ be used in so-called canonical XSL-cipher action on V_n, $n = m \cdot \kappa$. One round of this cipher consists of two transformations. The first transformation S is nonlinear, $S = (\underbrace{\pi, \ldots, \pi}_{\kappa})$, $\pi \in S_{V_m}$. The second transformation is a multiplication of vectors from V_n by the matrix Λ from the right. Round keys of the canonical XSL-cipher are equal to 0.

In the case of the canonical XSL-cipher a set $\mathfrak{W}_1$ is replaced by

$$\mathfrak{W}_1^{(0)} = \{\mathfrak{L} = ((l'_i, l''_i),\, i = 1, \ldots, k) | \, \forall i \in \{1, \ldots, k\} : (l'_i = 0 \Leftrightarrow l''_i = 0)\}.$$

By $C_\tau(\Lambda)$ we denote the linear medium of the canonical XSL-cipher with τ rounds. The branch number of matrix Λ is defined as $\rho_{1,2}(\Lambda) = \theta_1(C_2(\Lambda))$. If the column $l \in V_n^*$ is composed of columns $l_1, \ldots, l_\kappa \in V_m^*$ and

$$w(l) = \left|\{j \in \{1, \ldots, \kappa\} \middle| \, l_j \neq 0\}\right|$$

then we can see that

$$\rho_{1,2}(\Lambda) = \min_{l \in V_n^* \setminus \{0\}} \left(w(l) + w(\Lambda l)\right).$$

Indices 1 and 2 in the notation $\rho_{1,2}(\Lambda)$ correspond to $s = 1$ and $\tau = 2$.

Matrices $\Lambda \in GL(n, 2)$ having large branch number $\rho_{1,2}(\Lambda)$ are preferable from the viewpoint of a cryptographic design.

2.2. Let $P \in GL(n,2)$ be a substitution matrix, then $\rho_{1,2}(P) = 2$. For nonsingular matrices the smallest value of this characteristic is 2.

Thus the branch number $\rho_{1,2}(P)$ does not distinguish substitution matrices. This fact is a main disadvantage of $\rho_{1,2}(P)$.

A set of characteristics

$$\rho_{1,\tau}(P) = \theta_1(C_\tau(P)) = \tau,\ \tau \geqslant 2,$$

does not change the situation.

But distinguishing of different permutations $P \in S_n$ according to the degree of diffusion (in the sense of Shannon) takes place in a cryptographic practice. Originally the diffusion properties of permutations P are defined by the avalanche effect. Theorem 1 below describes optimal in this sense permutations P.

In order to formulate the Theorem 1 we must introduce some definitions.

The substitution

$$\pi : V_n \to V_n,\ x = (x_1, \ldots, x_n) \mapsto (y_1, \ldots, y_n) = y,$$

is called *significant* if the following conditions hold:

- y_j depends significantly on x_i for all $i, j \in \{1, \ldots, n\}$;
- x_i depends significantly on y_j for all i, j.

The SP-ciphers comprising significant substitutions are called *canonical*.

Other concept refers to directed graphs. If there exists only one directed path from i to j for any vertices $i, j \in \{1, \ldots, n\}$ and this path contains r edges then directed graph Γ on $n > 1$ vertices is called *∂-graph of order* $r \geqslant 1$.

For example, the de Bruijn graph on $n = m^r$ vertices is ∂-graph.

Further set $\{1, \ldots, n\}$ is divided into m-subsets

$$N(j) = \{(j-1)m + 1, (j-1)m + 2, \ldots, jm\},\ j = 1, \ldots, \kappa.$$

The substitutions π of transformation S act on vectors from V_m; the components of these vectors have numbers $N(j)$, $j = 1, \ldots, \kappa$.

The permutation $P : V_n \to V_n$ is associated with directed graph $\Gamma(P)$ on a set of vertices $\{1, \ldots, n\}$.

Theorem 1

If $n = m^r$ in the canonical SP-cipher and the graph $\Gamma(P)$ is ∂-graph of order r, then the substitution $(SP)^r : V_n \to V_n$ is significant.

2.3. The results of the previous sections mean that if we want to measure diffusion properties of matrices $\Lambda \in GL(m\kappa, 2)$ by the LMCD of canonical XSL-ciphers with τ rounds and linear medium C_τ then we must consider the LMCD $\theta_s(C_\tau)$ corresponding to s-dimensional linear cryptanalysis.

In this paper we propose a 2-parameter set of branch numbers $\rho_{s,\tau}(\Lambda)$, $s = 1, \ldots, m$, $\tau = 2, \ldots, \lceil \log_m n \rceil$, as a diffusion characteristic of matrix $\Lambda \in GL(m\kappa, 2)$.

We call this characteristic *the matrix's linear characteristic of diffusion* (MLCD). Branch numbers corresponding to small values of s, τ are more important.

We use branch numbers $\rho_{s,\tau}(\Lambda_1)$ and $\rho_{s,\tau}(\Lambda_2)$ to compare diffusion properties of matrices $\Lambda_1, \Lambda_2 \in GL(n, 2)$ only if we can not do it using $\rho_{s',\tau'}(\Lambda_1)$ and $\rho_{s',\tau'}(\Lambda_2)$, $s' \leqslant s$, $\tau' \leqslant \tau$, $(s', \tau') \neq (s, \tau)$.

Consider the conformal system

$$\mathfrak{L} = ((l'_i, l''_i), i = 1, \dots, \tau) =$$
$$= ((l'_{ij}, l''_{ij}),\ i = 1, \dots, \tau,\ j = 1, \dots, \kappa) \in \mathfrak{W}_s,$$

where $l''_i = \Lambda l'_{i+1}$, $i = 1, \dots, \tau - 1$.

We put

$$\theta_s(C_\tau) = \min_{\mathfrak{L} \in \mathfrak{W}_s^{(0)} \setminus \{0\}} \theta_{\mathfrak{L}},$$

where $\mathfrak{W}_s^{(0)} = \{\mathfrak{L} \in \mathfrak{W}_s \,\big|\, l'_{ij} = 0 \Leftrightarrow (\Lambda l'_{i+1})_j = 0,\ i = 1, \dots, \tau - 1,\ j = 1, \dots, \kappa\}$, $\theta_{\mathfrak{L}} = \left|\left\{(i,j) \in \{1, \dots, \tau\} \times \{1, \dots, \kappa\} \,\middle|\, l'_{ij} \neq 0\right\}\right|$.

Theorem 2

Suppose $n = m^r$, $P_j \in S_{m^r}$, $j = 0, 1, \ldots, r-2$,

$$P_j(i_0, i_1, \ldots, i_{r-j-1}, i_{r-j}, \ldots, i_{r-1}) = \\ = (i_1, \ldots, i_{r-j-1}, i_0, i_{r-j}, \ldots, i_{r-1}),$$

$i_0, i_1, \ldots, i_{r-1} \in \{0, 1, \ldots, m-1\}$. Then

$$\rho_{m,\tau}(P_j) = \tau m^{\tau-1} \text{ for } j + \tau \leq r$$

and

$$\rho_{m,\tau}(P_j) = \tau m^{r-1-j} \text{ for } j + \tau > r.$$

According to Theorem 2 we have

$$\rho_{m,\tau}(P_{r-1}) < \rho_{m,\tau}(P_{r-2}) < \ldots < \rho_{m,\tau}(P_{r-\tau+2}) <$$
$$< \rho_{m,\tau}(P_{r-\tau+1}) < \rho_{m,\tau}(P_{r-\tau}) = \rho_{m,\tau}(P_{r-\tau-1}) = \ldots =$$
$$= \rho_{m,\tau}(P_1) = \rho_{m,\tau}(P_0).$$

Characteristic $\rho_{m,\tau}$ does not distinguish permutations $P_0, P_1, \ldots, P_{r-\tau}$ for $\tau < r$ and this fact is a defect of $\rho_{m,\tau}$. But values $\rho_{m,\tau}(P_j)$, $j = 0, 1, \ldots, r-2, r-1$ are completely different for $\tau = r$, and the smaller j, the greater the value of this characteristic.

Thank you for attention!!