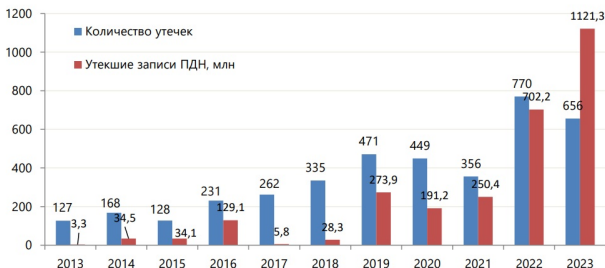


Asymptotic Distribution of the Correlation Coefficients of Linear Approximation

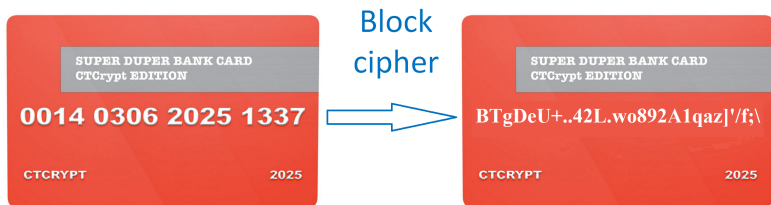
Denis Fomin, Aleksandr Kargin

- **Personal data** is transmitted through online services on a daily basis.
- As the volume of transmitted and stored information grows, so does the **need to protect it** from unauthorized access, leakage, and distortion.
- The rise in **data breaches** underscores the necessity for robust security measures¹.



¹ [InfoWatch](#). Россия: утечки информации ограниченного доступа, 2022–2023 годы. //

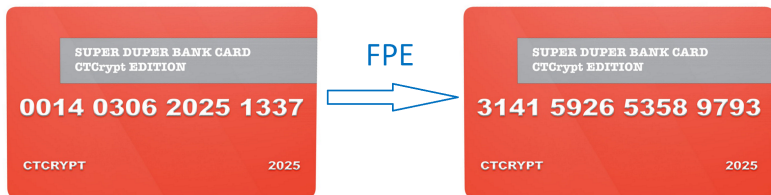
- Block ciphers are widely used in modern cryptographic systems; however, they **do not preserve the data structure**.



- Preserving the data format is critical in scenarios where **data structure is strictly regulated**, such as protecting payment details, identifiers, dates, and other structured fields.

FPE (Format-Preseving Encryption)

- Encrypted data keeps its original format.
- Compatible with existing systems.
- Reduce infrastructure upgrade costs.
- Already standardised in the US and South Korea.



- **FPE schemes based on the Feistel network** is the most widely used method.
- The standards FF1, FF3-1 (USA) and FEA-1, FEA-2 (South Korea) are built using **Feistel network**.

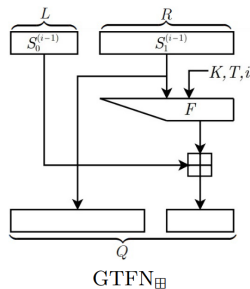
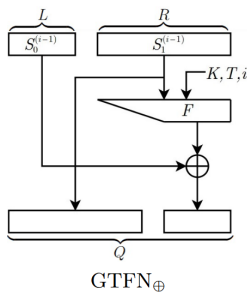


Рис.: Generalisation of the round transform design of FPEs built on an unbalanced Feistel network

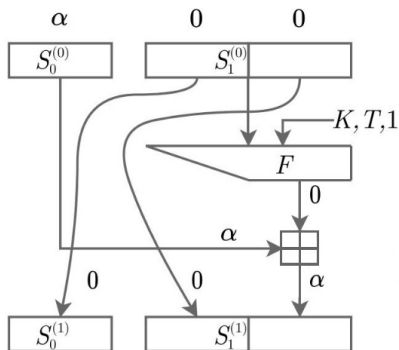
Using linear cryptanalysis², the insecurity of standardized algorithms FF1, FF3-1, FEA1, and FEA2 was demonstrated, and improved attacks³ on generic small-domain Feistel ciphers with alternating round tweaks were obtained.

Improved attacks on generic small-domain Feistel ciphers with alternating round tweaks have been achieved through linear cryptanalysis.

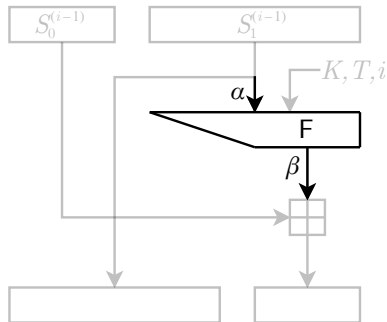
² [Beyne T.](#) Linear cryptanalysis of FF3-1 and FEA.

³ [Д. Б. Фомин.](#) Построение различителей на основе линейного и разностного методов анализа для сети Фейстеля.

The following local linear relations were used to construct consistent linear relations:



- In some standardized algorithms, the function F is defined based on the output of a block cipher.
- The problem of distinguishing the output of an FPE scheme from a random function was studied in⁴ and⁵.

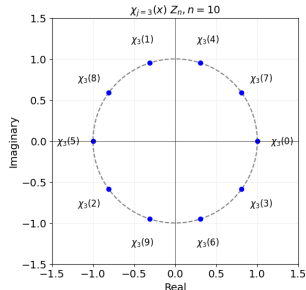


⁴ [Beyne T.](#) Linear cryptanalysis of FF3-1 and FEA.

⁵ [Д. Б. Фомин.](#) Построение различителей на основе линейного и разностного методов анализа

- A *character of an abelian group $\mathbf{X}$* is any homomorphism $\chi: \mathbf{X} \rightarrow \mathbb{C}^\times$ from the group $\mathbf{X}$ to the multiplicative group of the field of complex numbers $\mathbb{C}^\times$.
- For the cyclic group $\mathbb{Z}_n$, any character is a complex root of unity of degree n . All possible characters are described by n mappings of the form:

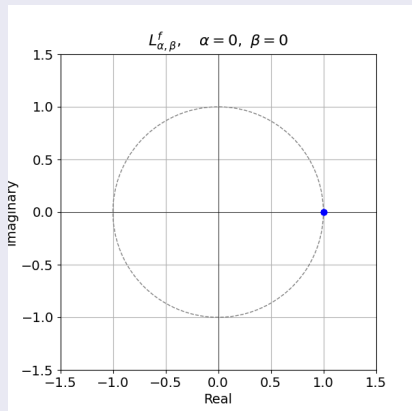
$$\chi_j(x) = \exp\left(\frac{2\pi i}{n} \cdot x \cdot j\right), \quad j \in \mathbb{Z}_n.$$



- The value

$$L_{\beta,\alpha}^f = \mathbb{M} \left[\psi_{\beta}(f(x)) \overline{\chi_{\alpha}(x)} \right]$$

is called, **the linear bias correlation coefficient**.



For binary spaces, the limit theorem for the distribution of the linear bias correlation coefficient is known⁶.

Theorem

Let $\mathbf{c}$ denote the correlation of a nontrivial linear approximation for a uniform random function $\mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$. The random variable $2^{n-1}(\mathbf{c} + 1)$ is binomially distributed with mean 2^{n-1} and variance 2^{n-2} . In particular, as $n \rightarrow \infty$, the distribution of $2^{n/2}\mathbf{c}$ converges to the standard normal distribution $\mathcal{N}(0, 1)$.

⁶ Daemen J., Rijmen V. Statistics of correlation and differentials in block ciphers.

A theorem from⁷, which is incorrect in the general case because it ignores important constraints related to the structure of groups and properties of characters.

Theorem 3.2. *Let G and H be finite Abelian groups and let $\mathbf{c}$ denote the correlation of a nontrivial linear approximation for a uniform random function $G \rightarrow H$ corresponding to non-real characters. The correlation $\mathbf{c}$ has mean zero and variance $1/|G|$. Furthermore, as $|G| \rightarrow \infty$, the distribution of $\sqrt{|G|}\mathbf{c}$ converges to the standard complex normal distribution $\mathcal{CN}(0, 1)$.*

Proof. Recall that a linear approximation corresponds to a pair of group characters $(\overline{\psi_1}, \psi_2)$. The random variable $\mathbf{c}$ can then be written as

$$\mathbf{c} = \frac{1}{|G|} \sum_{i=1}^{|G|} \psi_1(\mathbf{y}_i) \psi_2(x_i),$$

where $x_1, \dots, x_{|G|}$ are the elements of G and $\mathbf{y}_1, \dots, \mathbf{y}_{|G|}$ are independent uniform random variables on H . The mean of $\mathbf{c}$ is zero, since $\mathbb{E}\psi_1(\mathbf{y}_i) = 0$ by the orthogonality relations for group characters. In addition, it follows from $\mathbb{E}|\psi_1(\mathbf{y}_i)|^2 = 1$ that $\mathbb{E}|\mathbf{c}|^2 = 1/|G|$. Finally, the convergence to a normal distribution follows from the central limit theorem for the sum of independent identically distributed random variables. $\square$

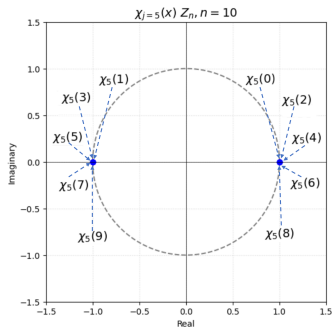
⁷ *Beyne T. Linear cryptanalysis of FF3-1 and FEA.*

Definition

Let $\mathbf{X}$ be an arbitrary abelian group, $|\mathbf{X}| \geq 2$, $\beta \in \mathbf{X} \setminus \{0\}$. Then the character ψ_β of the group $\mathbf{X}$ will be called a 2-character if for any $x \in \mathbf{X}$ the value of $\psi_\beta(x)$ can be represented as

$$\psi_\beta(x) = \exp \{ \pi i \cdot v(x) \}$$

for some $v(x) \in \mathbb{N}_0$.



Lemma

Let $\mathbf{X}$ and $\mathbf{Y}$ be arbitrary finite abelian groups satisfying the conditions $|\mathbf{Y}| \geq 2$ and $|\mathbf{X}| \geq 2$, and let $\beta \in \mathbf{X} \setminus \{0\}$ is a non-zero element of the group $\mathbf{Y}$ such that the character $\psi_\beta \in \widehat{\mathbf{Y}}$ is not a 2-character.

Consider a random variable Y uniformly distributed on the group $\mathbf{Y}$, that is $Y \sim \mathcal{U}(\mathbf{Y})$. For any fixed $x \in \mathbf{X}$, we define a complex-valued random variable $W = \psi_\beta(Y) \cdot \overline{\chi_\alpha(x)}$, where $\psi_\beta \in \widehat{\mathbf{Y}}$ and $\chi_\alpha \in \widehat{\mathbf{X}}$ are characters of the groups $\mathbf{X}$ and $\mathbf{Y}$ correspondingly. Let us represent W as $W = \Re(W) + i \cdot \Im(W)$, where $\Re(W)$ and $\Im(W)$ are the real and imaginary parts of W , respectively. Then

$$\mathbb{M}\Re(W) = \mathbb{M}\Im(W) = 0,$$

$$\mathbb{D}\Re(W) = \mathbb{D}\Im(W) = \frac{1}{2},$$

$$\text{cov}(\Re(W), \Im(W)) = 0.$$

Let's assume that:

- sequence of finite abelian groups $\{\mathbf{X}_n\}_{n=3}^{\infty}$ with corresponding dual character groups $\{\widehat{\mathbf{X}}_n\}_{n=3}^{\infty}$, $|\mathbf{X}_n| = n$ and the group $\mathbf{X}_n$ is not isomorphic to the group of the form $\mathbb{Z}_2 \dot{+} \mathbb{Z}_2 \dot{+} \dots \dot{+} \mathbb{Z}_2$;
- fixed finite abelian group $\mathbf{Y}$, where $|\mathbf{Y}| \geq 2$;
- for any $n \geq 3$:
 - $F_n : \mathbf{X}_n \rightarrow \mathbf{Y}$ — random function uniformly distributed on the set of all mappings $\mathfrak{F}(\mathbf{X}_n, \mathbf{Y})$ that is $F_n \sim \mathcal{U}(\mathfrak{F}(\mathbf{X}_n, \mathbf{Y}))$;
 - fixed elements $\alpha_n \in \mathbf{X}_n$ и $\beta_n \in \mathbf{Y} \setminus \{0\}$;
 - linear bias correlation coefficient, defined as

$$L_{\beta_n, \alpha_n}^{F_n} = \frac{1}{|\mathbf{X}_n|} \sum_{x \in \mathbf{X}_n} \psi_{\beta_n}(F_n(x)) \overline{\chi_{\alpha_n}(x)},$$

where $\psi_{\beta_n} \in \widehat{\mathbf{Y}}$ and $\chi_{\alpha_n} \in \widehat{\mathbf{X}}_n$ — characters of groups $\mathbf{Y}$ и $\mathbf{X}_n$, correspondingly.

Proposition

Suppose that for each $n \geq 3$ the character ψ_{β_n} is not a 2-character. Then for the sequence of normalized random variables

$$Z_n := \sqrt{|\mathbf{X}_n|} L_{\beta_n, \alpha_n}^{F_n}$$

has weak convergence:

$$Z_n \xrightarrow{d} Z \sim \mathcal{CN}(0, 1) \quad \text{at} \quad n \rightarrow \infty.$$

Let's note that $Z \sim \mathcal{CN}(0, 1)$, $\Re(Z) \sim \mathcal{N}(0, \frac{1}{2})$, $\Im(Z) \sim \mathcal{N}(0, \frac{1}{2})$

Corollary

Under the conditions of the statement 16, the following limit relations are true:

- 1** *The distribution of the random variable $|L_{\beta_n, \alpha_n}^{F_n}|$ converges to the Rayleigh distribution:*

$$|L_{\beta_n, \alpha_n}^{F_n}| \xrightarrow{d} \mathcal{R} \left(\frac{1}{\sqrt{2|\mathbf{X}_n|}} \right).$$

- 2** *The distribution of the random variable $LP_{\beta_n, \alpha_n}^{F_n} = |L_{\beta_n, \alpha_n}^{F_n}|^2$ converges to an exponential distribution:*

$$LP_{\beta_n, \alpha_n}^{F_n} \xrightarrow{d} \text{Exp} \left(\frac{1}{|\mathbf{X}_n|} \right).$$

ψ_β — 2-character

If ψ_β is 2-character, then the covariance matrix⁸ C^2 of the proposition 16 takes the following form:

$$C^2 = \text{diag} \left(\sum_{x \in \mathbf{X}} \left(\Re \left(\overline{\chi_\alpha(x)} \right) \right)^2, \sum_{x \in \mathbf{X}} \left(\Im \left(\overline{\chi_\alpha(x)} \right) \right)^2 \right).$$

Thus, convergence will be to a complex-valued random vector whose covariance matrix is given by the specified diagonal matrix.

⁸ V. Yu. Bentkus. A Lyapunov-type bound in R^d .

Cases where $Z \sim \mathcal{N}(0, 1)$

We can consider three (not necessarily mutually exclusive) cases where:

- 1 The group $\mathbf{X}$ is isomorphic to the group $\mathbb{Z}_2 \dot{+} \mathbb{Z}_2 \dot{+} \dots \dot{+} \mathbb{Z}_2$;
- 2 χ_α is 2-character;
- 3 $\alpha = 0$.

In all these cases, we obtain that the random variable Z asymptotically has a standard normal distribution when $|\mathbf{X}| \rightarrow \infty$. Note that the case when $\mathbf{X} = \mathbb{F}_2^n$, $\mathbf{Y} = \mathbb{F}_2^m$ is precisely a special case of the first case considered and convergence to a standard normally distributed random variable was obtained as early as in the paper⁹.

⁹ Daemen J., Rijmen V. Statistics of correlation and differentials in block ciphers.

- Clarification of the convergence conditions for linear bias correlation coefficients, taking into account the influence of 2-characters.
- The gaps present in the proofs of earlier studies are addressed and resolved.
- New perspectives are opened for the analysis of other cryptographic properties, such as the differential behaviour of functions.
- Of particular interest is the study of distributions not only of individual round functions, but also of sequences of such functions. .