UDC 004.056.55

ALGEBRAIC IMMUNITY OF SYMMETRIC CIPHERS

Aleksandr Kuznetsov[1], Roman Serhiienko[2], Dmytro Prokopovych-Tkachenko[3], Yuri Tarasenko[3], Ivan Belozertsev[1]

[1] V. N. Karazin Kharkiv National University, Svobody sq., 6, Kharkiv, 61022, Ukraine
kuznetsov@karazin.ua, ivanbelozersevv.jw@gmail.com

[2] National Army Academy named after Hetman Petro Sahaidachnyi, 32 Heroes of Maidan street, Lviv, 79012, Ukraine
romanserg69@gmail.com

[3] University of Customs and Finance, 2/4 Volodymyr Vernadsky str., Dnipro, 49000, Ukraine
omega2@email.dp.ua, me_dnepr@ua.fm

Reviewer: Anton Alekseychuk, Doctor of Sciences (Engineering), Full Prof., National Technical University of Ukraine "Kyiv Polytechnic Institute", Prosp. Peremohy, 37, Kyiv, 03056, Ukraine
alex-dtn@ukr.net

Received on November 2017

Abstract. *A key component of modern symmetric ciphers are nonlinear blocks (non-linear substitutions, substitution tables, S-boxes) that perform functions of hiding statistical links of plaintext and ciphertext, mixing and disseminating data, and introducing nonlinearity into the encryption procedure to counter various crypto-analytical and statistical attacks. The effectiveness of a symmetric cipher, its resistance to the majority of known cryptographic attacks and the level of information technology security provided by it directly depend on the performance of nonlinear nodes (balance, nonlinearity, autocorrelation, correlation immunity etc.). In this paper various methods for calculating algebraic immunity are examined, their interrelation is studied, and the results of comparative studies of the algebraic immunity of nonlinear blocks of the most well-known modern symmetric ciphers are presented.*

Keywords: *symmetric ciphers, algebraic immunity, nonlinear substitution blocks.*

1 Introduction

Cryptographic transformation plays an important role in ensuring the security of modern information systems and technologies [1, 2]. Symmetric ciphers because of their simplicity, efficiency and multifunctionality are used in almost all modern cryptographic protocols, and also as an integral part of other cryptographic primitives: hashing, pseudorandom sequence generation, password generation etc. Consequently, analysis and investigation of methods for synthesizing symmetric cryptographic primitives, the development and theoretical justification of criteria and performance indicators, including individual units of modern cyphers is important and relevant scientific and technical problem.

A key component of modern symmetric ciphers are nonlinear blocks (non-linear substitutions, substitution tables, S-boxes) that perform functions of hiding statistical links of plaintext and ciphertext, mixing and disseminating data, and introducing nonlinearity into the encryption procedure to counter various crypto-analytical and statistical attacks. Thus, the effectiveness of a symmetric cipher, its resistance to the majority of known cryptographic attacks and the level of information technology security provided by it directly depend on the performance of nonlinear nodes (balance, nonlinearity, autocorrelation, correlation immunity etc.).

Certain indices of the effectiveness of non-linear blocks of symmetric ciphers were considered in [3-9]. The concept of algebraic immunity was first introduced in [10,11] for estimating the stability of Boolean functions to the so-called algebraic cryptanalysis, proposed in [12]. In [13] these positions were generalized for Boolean mappings (S-blocks), to calculate algebraic immunity, the mathematical apparatus of Gröbner bases is used.

In this paper various methods for calculating algebraic immunity are examined, their interrelation is studied, and the results of comparative studies of the algebraic immunity of nonlinear blocks of the most well-known modern symmetric ciphers are presented.

2 Algebraic immunity of Boolean functions

The concept of algebraic immunity was first introduced in [10,11] and is considered in detail in the dissertation [14]. We introduce the definitions and notations necessary for the subsequent discussion, following the formulations adopted in [14].

Let $GF(2)$ be a binary field and $GF(2)^n$ – n-dimensional vector space over $GF(2)$.

Boolean function $f(x)$ of n variables is a mapping $f(x):GF(2)^n \to GF(2)$ where $x=(x_1,...,x_n)$.

Truth table of a Boolean function $f(x)$ of n variables is a binary output vector of the values of the function that contains 2^n elements, each element belongs to the set {0, 1}.

Algebraic normal form (Zhegalkin polynomial) of a Boolean function $f(x)$ of n variables is denoted in form:

$$f(x)=a_0 \oplus a_1x_1 \oplus a_2x_2 \oplus ... \oplus a_nx_n \oplus a_{12}x_1x_2 \oplus a_{13}x_1x_3 \oplus ... \oplus a_{(n-1)n}x_{n-1}x_n \oplus ... \oplus a_{123...n}x_1x_2x_3...x_n,$$

where the coefficients $a_i \in \{0,1\}$ and each Boolean function is implemented by the Zhegalkin polynomial uniquely, i.e. each representation of $f(x)$ corresponds to a unique truth table.

Algebraic degree $Deg(f)$ of a Boolean function $f(x)$ is a the number of variables in the longest term of the algebraic normal form of a function having a nonzero coefficient a_i. At the same time we consider $Deg(0)=0$.

Let’s denote as V_n the set of all mappings $GF(2)^n \to GF(2)$, i.e. this is the set of all possible Boolean functions $f(x)$ of n variables.

The set V_n we will consider both as the ring of Boolean functions and as a vector (linear) space over the binary field, i. e. $V_n = GF(2)^{2^n}$.

The Boolean function $g \in V_n$ is called the *annihilator* of a function $f \in V_n$, if $f \cdot g = 0$ or $(f+1)\cdot g = 0$.

The set of distinct annihilators of a Boolean function $g(x)$ forms a linear space, let’s denote it by $Ann(f)=\{g \in V_n \mid f \cdot g = 0\}$.

Let’s denote the linear space of annihilators of degree $\le d$ as

$$A_d^n(f)=\{g \in V_n \mid f \cdot g = 0, Deg(g) \le d\} \subset Ann(f).$$

The concept of annihilators of Boolean functions is closely related to the evaluation of the effectiveness of algebraic cryptanalysis of stream ciphers [10]. In particular, when using a filtering generator (see Fig. 1) of pseudo-random sequences (PRS) the search for the initial state of the linear feedback shift register (LFSR) is associated with a decrease in the degree of the joint system of polynomial Boolean equations.

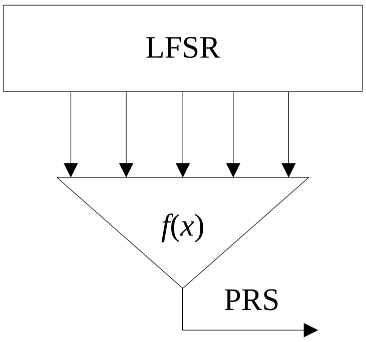

Fig. 1 – Block diagram of the filter generator PRS

Algorithm of algebraic cryptanalysis proposed in [10] allows under certain conditions, regarding the part of the intercepted output sequence (PRS), to find the initial state of the LFSR with time complexity $O\left((S_n^d)^3\right)$, where

$$S_n^d = \sum_{i=0}^{d} \frac{n!}{i!(n-i)!}$$

and d is the least degree of the non-zero annihilator of the filtering Boolean function $f(x)$ or its inversion $f(x)+1$.

Thus, the aim of algebraic cryptanalysis is the search for nonzero annihilators, or at least an estimation of their minimal degree. To this end, the definition of *algebraic immunity* $AI(f)$ of a Boolean function $f \in V_n$ was introduced in [11]:

$$AI(f) = \min\{D\text{eg}(g) \mid g \in Ann(f) \text{ or } g \in Ann(f+1)\}.$$

The value of $AI(f)$ is numerically equal to the minimal degree of such a Boolean function $g \in V_n$, that $f \cdot g = 0$ or $(f+1) \cdot g = 0$.

Using the concept of a linear space of annihilators of degree $\le d$ let's denote:

$$AI(f) = \min\{d \mid A_d^n(f) \ne 0 \,\text{или}\, A_d^n(f+1) \ne 0\}, \tag{1}$$

i.e. for evaluating the algebraic immunity of a Boolean function $f \in V_n$ it suffices to find a nonzero basis of the space of annihilators of the least degree of d.

The value d allows to quantify the complexity of algebraic cryptanalysis and, if sufficiently large d, to guarantee the resistance of a stream cryptographic algorithm to an algebraic attack.

Algorithm for computing the algebraic immunity of Boolean functions. One of the algorithms for calculating the algebraic immunity of Boolean functions is presented in the thesis [14]. It is based on the construction of a basis for the linear space of annihilators $A_d^n(f)$ of a given degree d. By increasing d iteratively and repeating the construction of the basis of the space $A_d^n(f)$, we obtain the $AI(f)$ estimation by the formula (1), i.e. through a nonzero basis of annihilators of the least degree.

It is necessary to introduce the following additional notation for description the essence of the algorithm.

Let's denote a monomial with respect to variables $x_1,...,x_n$ as

$$x^u = \prod_{i=1}^{n} x_i^{u_i} = \begin{cases} x_i, u_i = 1, \\ 1, u_i = 0, \end{cases}$$

where vectors $x, u \in V_2^n, x = (x_1,...,x_n), u = (u_1,...,u_n)$.

The degree of the monomial x^u is determined by the Hamming weight (the number of nonzero coordinates) $w_h(u)$ of the vector $u = (u_1,...,u_n)$, i.e.

$$Deg(x^u) = w_h(u).$$

Taking these notations into account, the Boolean function $f(x)$ in algebraic normal form (in the form of Zhegalkin polynomial) can be written in the form

$$f(x) = \sum_{u \in GF(2)^n} a_u x^u,\ a_u \in GF(2). \tag{2}$$

The function (annihilator) $g \in A_d^n(f)$ can also be represented it in the form of Zhegalkin polynomial

$$g(x)=\sum_{v\in GF(2)^n:w_h(v)\le d} b_v x^v\,, \tag{3}$$

where $b_v\in GF(2)$ – unknown annihilator coefficients, $w_h(v)$ – Hamming weight of the vector $v=(v_1,...,v_n)$. The function g belongs to the space $A_d^n(f)$ only if equality $f(x)\cdot g(x)=0$ holds for any $x\in GF(2)^n$.

By substituting (2) and (3) we obtain

$$f(x)\cdot g(x)=\left(\sum_{u\in GF(2)^n} a_u x^u\right)\left(\sum_{v\in GF(2)^n:w_h(v)\le d} b_v x^v\right)=\sum_{u\in GF(2)^n}\left(\sum_{v\in GF(2)^n:w_h(v)\le d} a_u b_v x^{u\vee v}\right)=0\,,$$

where $u\vee v=(u_1\vee v_1,...,u_n\vee v_n)$, $\vee$ – disjunction (logical OR operation).

After grouping the terms by the common factor, we obtain the equality:

$$\sum_{w\in GF(2)^n}\left(\sum_{a_u,b_v:a_u\vee b_v=w} a_u b_v\right)x^w=0\,, \tag{4}$$

which holds for any $w\in GF(2)^n$. Consequently, a system of linear homogeneous equations is obtained

$$\left\{\sum_{a_u,b_v:a_u\vee b_v=w} a_u b_v=0\,,\ \forall w\in GF(2)^n\right. \tag{5}$$

relatively unknown coefficients b_v of annihilation $g(x)$.

The solution of the system (5), for example, by the Gauss method, determines the basis of the space $A_d^n(f)$.

Pattern. For $n=2$ and $d=1$

$$f(x)=a_{00}+a_{10}x_1+a_{01}x_2+a_{11}x_1x_2\,,$$
$$g(x)=b_{00}+b_{10}x_1+b_{01}x_2\,.$$

After substitution $f(x)\cdot g(x)=0$ it follows

$$\begin{aligned}f(x)\cdot g(x)=a_{00}b_{00}+(a_{00}b_{10}+a_{10}b_{10}+a_{10}b_{00})x_1+(a_{00}b_{01}+a_{01}b_{01}+a_{01}b_{00})x_2+\\+(a_{10}b_{01}+a_{01}b_{10}+a_{11}b_{00}+a_{11}b_{10}+a_{11}b_{01})x_1x_2=0,\end{aligned}$$

from which it comes to a system of linear homogeneous equations:

$$\begin{cases}a_{00}b_{00}=0,\\ a_{00}b_{10}+a_{10}b_{10}+a_{10}b_{00}=0,\\ a_{00}b_{01}+a_{01}b_{01}+a_{01}b_{00}=0,\\ a_{10}b_{01}+a_{01}b_{10}+a_{11}b_{00}+a_{11}b_{10}+a_{11}b_{01}=0\end{cases}$$

relatively unknown b_{00},b_{10},b_{01} – coefficients of the function $g(x)$.

Then, for example, for the function $f(x)=x_1+x_2$ (i.e. for $a_{00}=a_{11}=0$ and $a_{10}=a_{01}=1$) we've got the system:

$$\begin{cases}b_{10}+b_{00}=0,\\ b_{01}+b_{00}=0,\\ b_{01}+b_{10}=0,\end{cases}$$

which satisfies only two solutions:

$$b_{00} = b_{10} = b_{01} = 0 \text{, i.e. } g(x) = 0,$$
$$b_{00} = b_{10} = b_{01} = 1 \text{, i.e. } g(x) = 1 + x_1 + x_2.$$

A close inspection shows that $g(x) = 1 + x_1 + x_2$ is indeed an annihilator of the function $f(x) = x_1 + x_2$:

$$f(x) \cdot g(x) = (x_1 + x_2)(1 + x_1 + x_2) = x_1 + x_2 + x_1 + x_1x_2 + x_1x_2 + x_2 = 0.$$

Summarizing the aforesaid, we define the basic steps of **the algorithm for finding the basis of the annihilator space** [14].

Input: $n \in \mathrm{N}, d \in \{1,...,n\}$, function $f(x)$ (given by a list of monomials x^u with nonzero coefficients a_u in (2)).

Output: Linear space $A_d^n(f)$ given in the form of a parametric family of Zhegalkin polynomials in n Boolean variables of degree $\le d$.

Step 1. Represent the functions $f(x)$ and $g(x)$ in the form of the sums (2) and (3), respectively.

Step 2. Expand the brackets in the product $f(x) \cdot g(x)$ and, by grouping the summands $a_u b_v x^w$ by sorting them by $a_u \vee b_v = w$, obtain the equation (4).

Step 3. Compose a system of linear homogeneous equations (5).

Step 4. Find the general solution of the system (5) in parametric form and feed it to the output of the algorithm.

The dissertation [14] gives an estimate $O\left(m \cdot \left(S_n^d\right)^3\right)$ of the bit complexity of the considered algorithm, where m is the number of non-zero coefficients a_u in (2).

Using the considered above algorithm for searching the basis of the annihilator space, we can calculate the algebraic immunity of a Boolean function $f(x)$ by sequentially scanning all the values $d > 0$ until we obtain a nonzero space of annihilators $A_d^n(f)$ or $A_d^n(f+1)$. The minimum value, for which $A_d^n(f) \ne 0$ and/or $A_d^n(f+1) \ne 0$, corresponds to the value of the algebraic immunity of a Boolean function $f(x)$.

Algorithm for calculating algebraic immunity $AI(f)$.

Input: $n \in \mathrm{N}$, function $f(x)$ (given by a list of monomials x^u with nonzero coefficients a_u in (2)).

Output: The value of Algebraic Immunity $AI(f)$.

Step 1. Assign $d = 1$.

Step 2. Calculate the space of annihilators $A_d^n(f)$ and $A_d^n(f+1)$.

Step 3. If $A_d^n(f) = 0$ and $A_d^n(f+1) = 0$ assign $d = d+1$ and go to step 2.

Step 4. If $A_d^n(f) \ne 0$ and/or $A_d^n(f+1) \ne 0$ assign $AI(f) = d$ and feed it to the output of the algorithm.

3 Algebraic immunity of Boolean mappings (S-boxes)

The concept of algebraic immunity of Boolean functions in [13] is generalized to the case of Boolean mappings $F : GF(2)^n \to GF(2)^n$ (*vector Boolean functions*), which are implemented by substitution blocks (substitution tables, S-boxes) of block symmetric ciphers. To determine the algebraic immunity $AI(F)$ we'll use the terms and definitions from [15].

Let state the natural numbers n, m, and the field K. Let consider a finite system S of m algebraic equations

$$\begin{cases} P_1(x_1, x_2, ..., x_n) = 0, \\ P_2(x_1, x_2, ..., x_n) = 0, \\ ... \\ P_m(x_1, x_2, ..., x_n) = 0 \end{cases} \tag{6}$$

of variables $x_1, x_2, ..., x_n$ with coefficients over the field K.

Let $K[x_1, x_2, ..., x_n]$ is the set of all polynomials in variables $x_1, x_2, ..., x_n$ with coefficients over the field K. On this set the operations of addition and multiplication are defined, and the set itself is called the *polynomial ring*. This ring is commutative (for any elements $a,b \in K[x_1, x_2, ..., x_n]$ holds the equality $a \cdot b = b \cdot a$), with an identity (for all $a \in K[x_1, x_2, ..., x_n]$ holds the equality $a \cdot e = a$, where $e = 1$).

A nonempty subset I of a commutative ring with identity R is called an *ideal* in R (denoted as $I \lhd R$) if the following two conditions are satisfied:

– for any elements $a,b \in I$ element $a - b \in I$;

– for any $a \in I$ и $c \in R$ element $a \cdot c \in R$.

Elements $a_1, a_2, ..., a_k$ constitute the *basis of the ideal*

$$I = (a_1, a_2, ..., a_k) = \{a_1 \cdot r_1 + a_2 \cdot r_2 + ... + a_k \cdot r_k; r_1, r_2, ..., r_k \in R\} \subseteq R.$$

It is said that an ideal $I \lhd R$ *admits a finite basis* if it contains elements $a_1, a_2, ..., a_k$ such that $I = (a_1, a_2, ..., a_k)$.

The fundamental ***Hilbert's basis theorem*** states that each ideal $I \lhd K[x_1, x_2, ..., x_n]$ admits a finite basis, i.e. there are such $f_1(x_1, x_2, ..., x_n)$, $f_2(x_1, x_2, ..., x_n)$, …, $f_k(x_1, x_2, ..., x_n) \in I$, that

$$I = (f_1, f_2, ..., f_k) = \{f_1 \cdot r_1 + f_2 \cdot r_2 + ... + f_k \cdot r_k; r_1, r_2, ..., r_k \in K[x_1, x_2, ..., x_n]\}.$$

Let associate with the system S (6) the ideal I, generated by the polynomials $P_1(x_1, x_2, ..., x_n)$, $P_2(x_1, x_2, ..., x_n)$, …, $P_m(x_1, x_2, ..., x_n)$, corresponding to the equations of the system:

$$I(S) = (P_1, P_2, ..., P_m) = \{P_1 \cdot r_1 + P_2 \cdot r_2 + ... + P_m \cdot r_m; r_1, r_2, ..., r_m \in K[x_1, x_2, ..., x_n]\}.$$

If $F \in I(S)$, then for each solution $(X_1, X_2, ..., X_n)$ of system (6) holds the equality

$$\begin{aligned} & F(X_1, X_2, ..., X_n) = \\ & = P_1(X_1, X_2, ..., X_n) \cdot r_1(X_1, X_2, ..., X_n) + P_2(X_1, X_2, ..., X_n) \cdot r_2(X_1, X_2, ..., X_n) + ... + \\ & + P_m(X_1, X_2, ..., X_n) \cdot r_m(X_1, X_2, ..., X_n) = \\ & = 0 \cdot r_1(X_1, X_2, ..., X_n) + 0 \cdot r_2(X_1, X_2, ..., X_n) + ... + 0 \cdot r_m(X_1, X_2, ..., X_n) = 0. \end{aligned}.$$

If $\{P_1, P_2, ..., P_m\}$ and $\{\overline{P}_1, \overline{P}_2, ..., \overline{P}_k\}$ both are two bases of the same ideal I, then the system of algebraic equations

$$\begin{cases} P_1(x_1, x_2, ..., x_n) = 0, \\ P_2(x_1, x_2, ..., x_n) = 0, \\ ... \\ P_m(x_1, x_2, ..., x_n) = 0, \end{cases} \qquad \begin{cases} \overline{P}_1(x_1, x_2, ..., x_n) = 0, \\ \overline{P}_2(x_1, x_2, ..., x_n) = 0, \\ ... \\ \overline{P}_k(x_1, x_2, ..., x_n) = 0 \end{cases}$$

are equivalent, that is the sets of their solutions coincide.

Consequently, the set of solutions of a system of algebraic equations is uniquely determined by the ideal of the system, and the various bases of the same ideal correspond to equivalent systems [15].

Suppose that there is a certain polynomial $h(x_1, x_2, ..., x_n) \in K[x_1, x_2, ..., x_n]$ and it is required in a finite number of steps to find out whether it belongs to an ideal $I \lhd K[x_1, x_2, ..., x_n]$ given by its basis $I = (f_1, f_2, ..., f_m)$. In other words, it is necessary to solve the so-called *problem of occurrence*: to find out whether there exist such polynomials $r_1(x_1, x_2, ..., x_n)$, $r_2(x_1, x_2, ..., x_n)$, …, $r_m(x_1, x_2, ..., x_n)$, that $h = f_1 \cdot r_1 + f_2 \cdot r_2 + ... + f_m \cdot r_m$ and $h \in I = (f_1, f_2, ..., f_m)$.

The problem of occurrence is solved by simplifying the expression for $h(x_1, x_2, ..., x_n)$ using so called *reduction of a polynomial*. Let's write the polynomial $h(x_1, x_2, ..., x_n)$ as the sum: $h = h_C + h_M$, where h_C – senior monomial, and h_M – the sum of the remaining monomials in h. Suppose also that h_C is divisible by the leading term f_{iC} of one of the polynomials f_i, i.e. $h_C = f_{iC} \cdot Q$ and $h = f_{iC} \cdot Q + h_M$ for some monomial Q. Then the *operation of reduction* is given by

$$h_1 = h - f_i \cdot Q = f_{iC} \cdot Q + h_M - f_{iC} \cdot Q - f_{iM} \cdot Q = h_M + (-f_{iM}) \cdot Q,$$

where f_{iM} - the sum of the remaining monomials in $f_i = f_{iC} + f_{iM}$. Herewith the leading term of the polynomial h_1 is less than the leading term of the polynomial h. If a polynomial h belongs to an ideal $I = (f_1, f_2, ..., f_m)$, then the reduced polynomial h_1 will also belong to this ideal. Indeed if $h \in (f_1, f_2, ..., f_m)$ then $h - h_1 = f_i Q \in (f_1, f_2, ..., f_m)$. Consequently, the problem of occurrence can now be solved no longer for a polynomial h, but for a reduced polynomial h_1. If for a finite number of reductions the polynomial h is reduced to zero (zero belongs to any ideal), then $h \in (f_1, f_2, ..., f_m)$.

Basis $f_1, f_2, ..., f_m$ of ideal $I = (f_1, f_2, ..., f_m)$ is called ***the Gröbner basis*** of this ideal if every polynomial $h \in I$ reduces to zero by means of $f_1, f_2, ..., f_m$. In other words the set of polynomials $f_1, f_2, ..., f_m$ is a Gröbner basis in the ideal $I = (f_1, f_2, ..., f_m)$ if for any $h \in I$ monomial h_C is divisible by one of the monomials $f_{1C}, f_{2C}, ..., f_{mC}$ [15].

For the operation of reduction of polynomials the concept of the leading term is used. In other words, it is assumed that on a set of all monomials of the ring $K[x_1, x_2, ..., x_n]$ the *linear order* (monomial ordering $\prec$) is given that satisfies the properties [16]:

– it follows from $x^u \prec x^v$ that $x^w \cdot x^u \prec x^w \cdot x^v$ for any monomials x^u, x^v, x^w (monomials are defined as (2), i.e. $x, u, v, w \in V_2^n, x = (x_1, ..., x_n), u = (u_1, ..., u_n), v = (v_1, ..., v_n), w = (w_1, ..., w_n)$);

– $1 \preceq x^v$ for any monomial x^u.

Some examples of monomial ordering are cited below:

– *dictionary* or *lexicographic order (lex)*: $x^u \prec_{\text{lex}} x^v$, if such i exists that $u_i < v_i$ and $u_j = v_j$ for $j < i$ (first the variables in monomials in the alphabetical order are ordered, and then the first difference in monomials is found);

– *degree lexicographic order (deglex)*: $x^u \prec_{\text{deglex}} x^v$, if $w_h(u) < w_h(v)$ or $w_h(u) = w_h(v)$, but with that $x^u \prec x^v$ in the alphabetical order (ordered by the sum of powers, in the case of equality of sums – by alphabetical order);

– *degree reverse lexicographic order (degrevlex)*: $x^u \prec_{\text{degrevlex}} x^v$, if $w_h(u) < w_h(v)$ or $w_h(u) = w_h(v)$, but with that $x^u \succ_{\text{lex}} x^v$ in the alphabetical order (ordered by the sum of powers, in the case of equality of sums – by reverse alphabetical order).

The solution of the problem of occurrence, i.e. the ascertainment of membership of a polynomial h to an ideal $I=(f_1, f_2,..., f_m)$, consists in constructing all possible reductions h by means of elements of the Gröbner basis of the ideal I. A polynomial h belongs to an ideal $I=(f_1, f_2,..., f_m)$ if and only if a zero is obtained as a result of reduction [15].

For each ideal $I \lhd K[x_1, x_2,..., x_n]$ there exists a Gröbner basis, and the construction of the Gröbner basis itself is based on the resolving the linkage [15]. The polynomials f_i and f_j have a linkage if their leading terms are both divisible by a non-constant monomial ω. Let $f_{iC}=\omega\cdot q_1$, $f_{jC}=\omega\cdot q_2$, where ω – the greatest common divisor of leading terms f_{iC} and f_{jC}. Let's consider the monomial $F_{i,j}=f_i\cdot q_2 - f_j\cdot q_1 \in I$ and reduce it using a basis $f_1, f_2,..., f_m$ as long as possible. If the resulting polynomial $F'_{i,j}\equiv 0$, then they say the linkage is solvable. Otherwise, the resulting polynomial $f_{m+1}=F'_{i,j}$ should be added to the basis $f_1, f_2,..., f_m$ of the ideal I after which the procedure for finding and reducing of linkage will be continued. After reducing the finite number of linkages a set $f_1, f_2,..., f_m, f_{m+1},..., f_M$ is obtained in which every linkage is solvable.

In accordance with the ***diamond lemma***, the basis $f_1, f_2,..., f_m$ of an ideal $I \lhd K[x_1, x_2,..., x_n]$ is a Gröbner basis only if there are no unsolvable linkages in it [15].

The resolving of the linkage allows to define the effective algorithm for constructing the Gröbner basis of the ideal $I=(f_1, f_2,..., f_m)$ (***Buchberger's algorithm***).

Step 1. Check whether the linkage in the set $f_1, f_2,..., f_m$ exists. If there are no linkages, then the set $f_1, f_2,..., f_m$ is a Gröbner basis of the ideal $I=(f_1, f_2,..., f_m)$. If linkages exist then go to step 2.

Step 2. Form a polynomial $F_{i,j}=f_i\cdot q_2 - f_j\cdot q_1$ with linkage of the polynomials f_i and f_j found in previous step and reduce it by means of a set $f_1, f_2,..., f_m$ as long as this is possible. If the polynomial is reduced to a nonzero polynomial f_{m+1} go to step 3, otherwise go to step 4.

Step 3. Add the polynomial f_{m+1} to the set $f_1, f_2,..., f_m$ and go to step 4.

Step 4. Pick up linkage didn't examined previously and go to step 2. If all the linkages are processed, then we derive the resulting set $f_1, f_2,..., f_m, f_{m+1},..., f_M$ in which all the linkages are solvable. This is the Gröbner basis of ideal $I=(f_1, f_2,..., f_m)$.

To date, other algorithms for constructing the Gröbner basis are known, for example algorithms F4, F5 [17,18]. The Gröbner basis can be simplified in the following methods [15].

1. *Minimization of the Gröbner basis*. If f_i and f_j are two elements of the Gröbner basis, with their leading terms f_{iC} and f_{jC} that are divisible by each other, for example, $f_{jC} \mid f_{iC}$, then the polynomial f_i can be removed from the set $f_1, f_2,..., f_m$. The Gröbner basis is called *minimal* if f_{iC} it is not divisible by f_{jC} for all $i \neq j$.
2. *Reduction of the Gröbner basis*. If some member q of the polynomial f_i is divisible by the leading term of the polynomial f_j, then we reduce q it with f_j and use the result of reduction to replace the term q in the polynomial f_i. In this case the Gröbner basis remains a Gröbner basis, the number of elements of the basis does not change, however the degrees of the polynomials $f_1, f_2,..., f_m$ decrease. The Gröbner basis is said to be *reduced* if no member of the polynomial f_i is divisible by the leading term of the polynomial f_j for all $i \neq j$.

The minimal reduced Gröbner basis of the ideal $I \lhd K[x_1, x_2,..., x_n]$ is uniquely defined (with unit coefficients at the highest powers of the basis elements), that is, it doesn't depend on the initial basis of the ideal $I=(f_1, f_2,..., f_m)$ and on the sequence of operations performed (but depends on the ordering of the variables $x_1, x_2, ..., x_n$) [15]. The concept of a minimal reduced Gröbner basis is used in the work of Jean-Charles Faugère [13] to determine the algebraic immunity of S-blocks (nonlinear complication nodes) of block symmetric ciphers. Let consider a non-linear block (S-box) of the block symmetric cipher (see Fig. 2), which implements the Boolean mapping $S: GF(2)^n \to GF(2)^m$ [1-9].

S-box is defined by a system of algebraic equations over a binary field:

$$\begin{cases} f_1(x_1, x_2, ..., x_n) = y_1, \\ f_2(x_1, x_2, ..., x_n) = y_2, \\ ... \\ f_m(x_1, x_2, ..., x_n) = y_m, \end{cases} \tag{7}$$

i.e. a collective of Boolean polynomials

$$\begin{gathered} y_1 - f_1(x_1, x_2, ..., x_n), \\ y_2 - f_2(x_1, x_2, ..., x_n), \\ \dots, \\ y_m - f_m(x_1, x_2, ..., x_n) \end{gathered} \tag{8}$$

in the ring $K[x_1, x_2, ..., x_n, y_1, y_2, ..., y_m]$ of variables $x_1, x_2, ..., x_n, y_1, y_2, ..., y_m$ with coefficients over the field $K = GF(2)$.

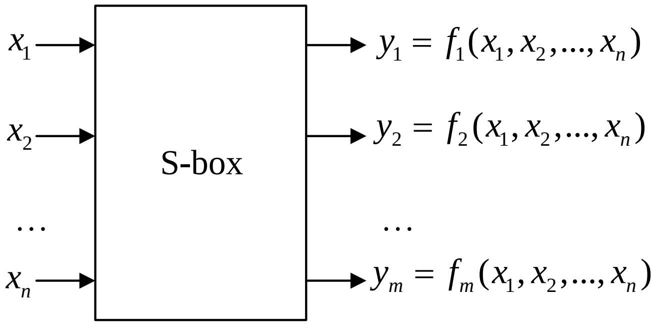
x_1
x_2
$\dots$
x_n
S-box
$y_1 = f_1(x_1, x_2, ..., x_n)$
$y_2 = f_2(x_1, x_2, ..., x_n)$
$\dots$
$y_m = f_m(x_1, x_2, ..., x_n)$

Fig. 2 – Block diagram of a non-linear block of a block symmetric cipher

With the system of equations (7), algebraically defining the structure of an S-block, we associate the ideal I generated by the polynomials (8):

$$\begin{aligned} I(S) &= (y_1 - f_1(x_1, x_2, ..., x_n), y_2 - f_1(x_1, x_2, ..., x_n), ..., y_m - f_m(x_1, x_2, ..., x_n)) = \\ &= \{(y_1 - f_1) \cdot r_1 + (y_2 - f_1) \cdot r_2 + ... + (y_m - f_m) \cdot r_m; r_1, r_2, ..., r_m \in GF(2)[x_1, x_2, ..., x_n, y_1, y_2, ..., y_m]\}. \end{aligned}$$

Algebraic immunity of a non-linear block of a block symmetric cipher is defined as the minimal degree of a polynomial P in an ideal $I(S)$ [13]:

$$AI(S) = \min\{\deg(P), P \in I(S) \triangleleft GF(2)[x_1, x_2, ..., x_n, y_1, y_2, ..., y_m]\}, \tag{9}$$

and the minimal reduced Gröbner basis of the ideal $I(S)$ for a degree reverse lexicographic order (*degrevlex*) contains a linear basis of polynomials P in $I(S)$, such that $AI(S) = \deg(P)$. In other words, to calculate algebraic immunity $AI(S)$ it is sufficient to construct a minimal reduced Gröbner basis of the ideal $I(S)$ given by equations (8) and to find a polynomial of minimal degree among the elements of this basis. The value of the minimum degree is the value of the algebraic immunity $AI(S)$ of the block symmetric cipher substitution box (S-box).

The link between the algebraic immunity of the S-block (9) and the Boolean function (1) is shown in [19, p. 337]. Consider a Boolean function $f_S(x_1,x_2,...,x_n,y_1,y_2,...,y_m):GF(2)^{2n}\to GF(2)$ whose values are defined as follows:

$$f_S(x_1,x_2,...,x_n,y_1,y_2,...,y_m)=\begin{cases}1,\forall i,\,j: f_i(x_1,x_2,...,x_n)=y_j,\\ 0,\exists i,\,j: f_i(x_1,x_2,...,x_n)\neq y_j.\end{cases}$$

The set of solutions of equation

$$f_S(x_1,x_2,...,x_n,y_1,y_2,...,y_m)-1=0$$

coincides with the set of solutions of system (7). Consequently, there are different bases (f_S-1) and $(y_1-f_1,y_2-f_1,...,y_m-f_m)$ of one ideal of equivalent systems, i.e.

$$I(f_S-1)=I(y_1-f_1,y_2-f_1,...,y_m-f_m).$$

Ideal of the space of annihilators $Ann(f_S)$ in the ring $GF(2)[x_1,x_2,...,x_n,y_1,y_2,...,y_m]$ coincides with the ideal $I(f_S-1)$, hence, the algebraic immunity (9) of the Boolean mapping $S:GF(2)^n\to GF(2)^m$ coincides with the minimal degree of nonzero polynomials belonging to the annihilator of the function f_S:

$$AI(S)=\min\{D\text{eg}(g)\,|\,g\in Ann(f_S)\}.$$

Thus, any S-block can be unambiguously described by a Boolean function [19], the algebraic immunity of this function can be calculated, for example, using the algorithm of paragraph 2.

4 Values of algebraic immunity of nonlinear blocks of modern ciphers

In this paper comparative studies of the algebraic immunity of nonlinear blocks of modern symmetric ciphers have been carried out. As objects of research, well-known and standardized on the national and/or international level block symmetric crypto-transformations are chosen:

- cryptographic algorithm AES, standardized in the US as a federal data processing standard FIPS-197 [20], and also internationally as a block cipher in ISO / IEC 18033-3 [21];
- cryptographic algorithm Camellia, standardized internationally as a block cipher in ISO/IEC 18033-3 [21];
- cryptographic algorithm CAST, standardized internationally as a block cipher in ISO/IEC 18033-3 [21];
- cryptographic algorithm SEED, standardized internationally as a block cipher in ISO/IEC 18033-3 [21];
- cryptographic algorithm “Kalyna”, national standard of Ukraine DSTU 7624:2014 [22];
- cryptographic algorithm “Kuznechik”, standardized in Russia as GOST 34.12-2015 [23];
- algorithm of symmetric encryption and integrity control “BelT”, the Republic of Belarus, standardized in STB 34.101.31-2011 [24];
- cryptographic hash function Whirlpool, based on block symmetric crypto-transformations, standardized internationally in ISO/IEC 10118-3:2004 [25].

To calculate algebraic immunity the expression (9) was used. For immediate calculations, the Magma software package [26] is used, which implements a wide range of functions related to algebra, group theory, rings and fields, number theory and many other branches of mathematics.

The tested blocks of the replacements, except for the S-block of the hash function of Whirlpool, were considered in detail in work [9], table 1 shows some results of the research.

The following notations [9] are used in the table:

- B – balance;
- N – non-linearity;

- A – autocorrelation;
- AD – algebraic degree;
- PC –propagation criterion;
- CI – correlation immunity.

Table 1 – Cryptographic properties of non-linear blocks of block ciphers

	B	**N**	**A**	**AD**	**PC**	**CI**	**AI**
AES	+	112	32	7	0	0	2
SEED	–	110	40	7	0	0	2
CAST-128	–	120	0	4	8	0	2
“Camellia”	+	112	32	7	0	0	2
“Kalina”	+	104	72	7	0	0	3
“Kuznechik”	+	102	72	7	0	0	3
“BelT”	+	104	72	7	0	0	3
“Whirlpool”	+	95	80	7	0	0	3

In the last column “AI” of Table 1 the values of the algebraic immunity of nonlinear substitution blocks of modern ciphers are listed. These data are obtained from (9) by constructing Gröbner bases of ideals $I(S)$ given by sets of polynomials (8) from equations (7) of the corresponding S-blocks.

The results obtained are indicative of the insufficient algebraic immunity of nonlinear boxes of block ciphers, which were developed in the late 90s – early 2000s. The algorithms considered (AES, SEED, CAST-128, “Camellia”), represented in the modern international standard ISO / IEC 18033-3, have relatively low algebraic immunity and can potentially be considered as real targets for constructing effective algebraic attacks.

Block symmetric crypto algorithms “Kalyna”, “Kuznechik”, “BelT”, as well as cryptographic function of hashing of Whirlpool, are developed taking into account the possible algebraic attacks. Nonlinear substitution blocks of these algorithms have high algebraic immunity and, apparently, will remain resistant to new methods of algebraic cryptanalysis.

5 Conclusion

Methods of algebraic cryptanalysis since early publications [27,28] have turned from abstract and inapplicable mathematical ideas into a developed section of modern cryptology that is widely discussed in the scientific community. To date, a huge number of research projects have been carried out in this field of knowledge, and obviously, in the coming years, effective algorithms for algebraic cryptanalysis of modern symmetric ciphers should appear.

In this paper some aspects of algebraic cryptanalysis were considered, in particular, methods for calculating the algebraic immunity of non-linear blocks of symmetric ciphers were studied. This concept first was introduced for stream cryptoalgorithms in [10,11], and was generalized in [13] to Boolean mappings, i.e. nonlinear blocks with arbitrary dimension of inputs and outputs. Algebraic immunity, in some sense, characterizes the complexity of solving a system of equations describing a non-linear block and thus allows one to obtain an idea of the resistance of a symmetric cipher to algebraic cryptanalysis. In particular, the algorithm of algebraic cryptanalysis of stream ciphers with filter-generator scheme was proposed in [10]. Complexity of implementing this algorithm is a function of the value of algebraic immunity of a cryptographic Boolean function.

The calculation of the algebraic immunity of a nonlinear block in the general case is associated with the construction of the Gröbner basis of the ideal of the polynomial ring given by polynomials from the equations of the permutation block. This problem is solved by computationally effective algorithms of Buchberger, F4, F5, etc. [15-18]. Moreover, the considered mathematical methods can also be used to search for effective algebraic attacks [19], which confirms the perspective and relevance of ongoing research in this field.

In this paper the algebraic immunity values substitution boxes of some modern ciphers are given. In particular, it was found out that the cryptoalgorithms developed at the end of the 90s – the beginning of the 2000s do not have the ultimate values of algebraic immunity, i.e. can be considered as targets for potential effective algebraic attacks. Block ciphers of the latest generation ("Kalina", "Kuznechik", "BelT") were developed taking into account the possible application of algebraic cryptanalysis and have uttermost values of algebraic immunity.

A promising direction is further research on methods of algebraic cryptanalysis, in particular, the use of quantum computing technologies to solve systems of algebraic equations that describe a symmetric cipher. According to the authors of this work, in this direction of research the most significant and interesting scientific results are expected.

References

[1] Alfred J. Menezes, Paul C. van Oorschot, Scott A. Vanstone. Handbook of Applied Cryptography – CRC Press, 1997. – 794 p.

[2] Gorbenko I.D., Gorbenko Y.I. Applied cryptology. Theory. Praxis. Exploitation: Textbook for higher educational institutions. – Kharkiv: Publishing house “Fort”, 2013. – 880 p.

[3] Bart Preneel. Analysis and Design of Cryptographic Hash Functions. [Electronic resource] – Access mode: homes.esat.kuleuven.be/~preneel/phd_preneel_feb1993.pdf.

[4] Carlet C. Vectorial Boolean functions for // Cambridge Univ. Press, Cambridge. – 95 p. [Electronic resource] – Access mode: www.math.univ-paris13.fr/~carlet/chap-vectorial-fcts-corr.pdf.

[5] Carlet C. Boolean functions for cryptography and error correcting codes // Cambridge Univ. Press, Cambridge. – 2007. – 148 p. [Electronic resource] – Access mode: www1.spms.ntu.edu.sg/~kkhoongm/chap-fcts-Bool.pdf.

[6] Zhuo Zepeng, Zhang Weiguo On correlation properties of Boolean functions // Chinese Journal of Electronics. Jan, Vol.20, 2011, №1, 143-146 pp.

[7] O’Connor L. An analysis of a class of algorithms for S-box construction // J. Cryptology. -1994. – p. 133-151.

[8] Clark J.A., Jacob J.L., Stepney S. The Design of S-Boxes by Simulated Annealing // New Generation Computing. – 2005. – 23(3). – p.219–231.

[9] Kuznetsov A.A., Bielozertsev I.N., Andrushkevich A.V. Analysis and comparative studies of nonlinear substitution blocks of modern block symmetric ciphers // Applied electronics. – Kharkiv: KhNURE. – 2015. – Vol. 14. №4. – p. 343 – 350.

[10] Courtois N., Meier W. Algebraic Attacks on Stream Ciphers with Linear Feedback, Eurocrypt 2003, LNCS 2656, Springer, 2003. – pp. 345-359.

[11] Meier W., Pasalic E., Carlet C: Algebraic Attacks and Decomposition of Boolean Functions, Eurocrypt 2004, LNCS 3027, Springer, 2004. – pp. 474-491.

[12] Nicolas Courtois; Josef Pieprzyk (2002). "Cryptanalysis of Block Ciphers with Overdefined Systems of Equations". LNCS. 2501: 267–287.

[13] Gw´enol´e Ars, Jean-Charles Faug`ere. Algebraic Immunities of functions over finite fields. [Research Report] RR-5532, INRIA. 2005, pp.17.

[14] Bayev Vladimir Valerievich. Effective algorithms for obtaining estimates of the algebraic immunity of Boolean functions: a thesis for the degree of candidate of physical and mathematical sciences: 01.01.09 / Bayev Vladimir Valerievich; [Institution of defense of a thesis: Moskow state university]. - Moskow, 2008. - 101 p.

[15] I.V. Arzhantsev. Gröbner bases and systems of algebraic equations. Summer school. Modern mathematics. Dubna, Yuly 2002. – Moskow: MCNMO, 2003. – 68 p.

[16] AI Zlobin, O. Sokolova. Computer algebra in the Sage system. Tutorial. - Moscow: MSTU named after Bauman, 2011. – 55 p.

[17] Faugère, J.-C. (June 1999). A new efficient algorithm for computing Gröbner bases (F4). Journal of Pure and Applied Algebra. Elsevier Science. 139 (1): 61–88.

[18] Faugère, J.-C. (July 2002). A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). Proceedings of the 2002 international symposium on Symbolic and algebraic computation (ISSAC). ACM Press: 75–83.

[19] Massimiliano Sala, Teo Mora, Ludovic Perret, Shojiro Sakata, Carlo Traverso Gröbner Bases, Coding, and Cryptography. Springer-Verlag Berlin Heidelberg. – 426 p.

[20] FIPS 197. National Institute of Standards and Technology. [Electronic resource]: Advanced Encryption Standard. – 2001. – Available at: http://www.nist.gov/aes.

[21] ISO/IEC 18033-3. Information technology – Security techniques – Encryption algorithms, Part 3: Block ciphers, 80 p.

[22] DSTU 7624:2014. Information technology. Cryptographic security of information. The algorithm of symmetric block transformation. – Kyiv: Ministry of economic development of Ukraine, 2015. – 238 p.

[23] GOST R 34.12-2015. Information technology. Cryptographic security of information. Block ciphers. – Moscow: Standartinform, 2015. – 25p.

[24] STB 34.101.31-2011. Information technology and security. Cryptographic algorithms for encryption and integrity control. – Minsk: Gosstandart, 2011. – 32 p.

[25] ISO/IEC 10118-3:2004. Information technology – Security techniques – Hash-functions – Part 3: Dedicated hash-functions, 94 p.

[26] Magma Computational Algebra System. Available at: http://magma.maths.usyd.edu.au/magma.

[27] Nicolas Courtois, Alexander Klimov, Jacques Patarin, Adi Shamir. Efficient Algorithms for Solving Overdefined Systems of Multivariate Polynomial Equations. Proceeding EUROCRYPT'00 Proceedings of the 19th international conference on Theory and application of cryptographic techniques. P. 392-407.

[28] Nicolas Courtois, Josef Pieprzyk. Cryptanalysis of Block Ciphers with Overdefined Systems of Equations. Advances in cryptology – ASIACRYPT 2002. P.267-287.

[29] Andrey Pyshkin. Algebraic Cryptanalysis of Block Ciphers Using Grobner Bases. Dissertation zur Erlangung des Grades Doktor rerum naturalium. Technischen Universit¨at Darmstadt. – Darmstadt, 2008, 118 p.

Рецензент: Антон Олексійчук, д.т.н., доцент, Інститут спеціального зв'язку та захисту інформації національного технічного університету України «КПІ», пр. Перемоги, 37, г. Киев, 03056, Україна.
E-mail: alex-dtn@ukr.net

Надійшло: Листопад 2017.

Автори:

Олександр Кузнецов, д.т.н., проф., академік Академії наук прикладної радіоелектроніки, Харківський національний університет ім. В. Н. Каразіна, пл. Свободи, 6, Харків, 61022, Україна. E-mail: kuznetsov@karazin.ua

Роман Сергієнко, Національна академія сухопутних військ імені гетьмана Петра Сагайдачного, вул. Героїв Майдану, 32, м. Львів, 79026, Україна. E-mail: romanserg69@gmail.com

Дмитро Прокопович-Ткаченко, Університет митної справи та фінансів, вул. Володимира Вернадського, 2/4, м. Дніпро, 49000, Україна. E-mail: omega2@email.dp.ua

Юрій Тарасенко, Університет митної справи та фінансів, вул. Володимира Вернадського, 2/4, м. Дніпро, 49000, Україна.
E-mail: me_dnepr@ua.fm

Іван Білозерцев, студент, Харківський національний університет імені В.Н. Каразіна, пл. Свободи, 6, Харків, 61022, Україна.
E-mail: ivanbelozersevv.jw@gmail.com

Алгебраїчний імунітет симетричних шифрів.

Анотація. Ключовим компонентом сучасних симетричних шифрів є нелінійні вузли (нелінійні підстановки, таблиці замін, S-блоки), які виконують функції приховування статистичних зв'язків відкритого тексту і шифртекста, перемішування і розсіювання даних, внесення нелінійності в процедуру шифрування для протистояння різним криптоаналітичних і статистичними атакам. Від показників ефективності нелінійних вузлів (збалансованості, нелінійності, автокореляції, кореляційної імунності та ін.) безпосередньо залежить ефективність симетричного шифру, його стійкість до більшості відомих криптографічних атак і рівень забезпечуваної безпеки інформаційних технологій. У даній роботі розглядаються різні методи розрахунку алгебраїчного імунітету, вивчається їх взаємозв'язок і наводяться результати порівняльних досліджень алгебраїчної імунності нелінійних вузлів найбільш відомих сучасних симетричних шифрів.

Ключові слова: симетричні шифри, алгебраїчний імунітет, нелінійні блоки підстановки.

Рецензент: Антон Алексейчук, д.т.н., доцент, Институт специальной связи и защиты информации национального технического университета Украины «КПИ», пр. Победы, 37, г. Киев, 03056, Украина.
E-mail: alex-dtn@ukr.net

Поступила: Ноябрь 2017.

Авторы:

Александр Кузнецов, д.т.н., проф., академик Академии наук прикладной радиоэлектроники, Харьковский национальный университет им. В. Н. Каразина, пл. Свободы, 6, Харьков, 61022, Украина. E-mail: kuznetsov@karazin.ua

Роман Сергиенко, Национальная академия сухопутных войск имени гетмана Петра Сагайдачного, ул. Героев Майдана, 32, г. Львов, 79026, Украина. E-mail: romanserg69@gmail.com

Дмитрий Прокопович-Ткаченко, Университет таможенного дела и финансов, ул. Владимира Вернадского, 2/4, г. Днепр, 49000, Украина. E-mail: omega2@email.dp.ua

Юрий Тарасенко, Университет таможенного дела и финансов, ул. Владимира Вернадского, 2/4, г. Днепр, 49000, Украина.
E-mail: me_dnepr@ua.fm

Иван Белозерцев, студент, Харьковский национальный университет имени В.Н. Каразина, пл. Свободы, 6, Харьков, 61022, Украина. E-mail: ivanbelozersevv.jw@gmail.com

Алгебраический иммунитет симметричных шифров.

Аннотация. Ключевым компонентом современных симметричных шифров являются нелинейные узлы (нелинейные подстановки, таблицы замен, S-блоки), которые выполняют функции скрытия статистических связей открытого текста и шифртекста, перемешивания и рассеивания данных, внесения нелинейности в процедуру зашифрования для противостояния различным криптоаналитическим и статистическим атакам. От показателей эффективности нелинейных узлов (сбалансированности, нелинейности, автокорреляции, корреляционной иммунности и пр.) непосредственно зависят эффективность симметричного шифра, его устойчивость к большинству известных криптографических атак и уровень обеспечиваемой им безопасности информационных технологий. В данной работе рассматриваются различные методы расчета алгебраического иммунитета, изучается их взаимосвязь и приводятся результаты сравнительных исследований алгебраической иммунности нелинейных узлов наиболее известных современных симметричных шифров.

Ключевые слова: симметричные шифры, алгебраический иммунитет, нелинейные блоки подстановки.