УДК 004.056.5

А.А. КУЗНЕЦОВ, д-р техн. наук, И.В. МОСКОВЧЕНКО, канд. техн. наук,
Д.И. ПРОКОПОВИЧ-ТКАЧЕНКО, канд. техн. наук, Т.Ю. КУЗНЕЦОВА

ЭВРИСТИЧЕСКИЕ МЕТОДЫ ГРАДИЕНТНОГО ПОИСКА КРИПТОГРАФИЧЕСКИХ БУЛЕВЫХ ФУНКЦИЙ

Введение

Важным элементом большинства современных симметричных шифров являются нелинейные блоки замен (S-блоков) [1 – 4], которые описываются с помощью булевых или, в общем случае, векторных криптографических функций [5 – 29]. Показатели стойкости таких функций (сбалансированность, нелинейность, автокорреляция и пр.) непосредственно влияют на эффективность симметричных шифров, их устойчивость к большинству современных криптоаналитических атак [5 – 17]. В частности, в работах [5 – 7] исследованы алгебраические свойства S-блоков современных блочных шифров, показано их влияние на устойчивость к алгебраическому криптоанализу. В [8 – 11] исследованы комбинаторные свойства нелинейных узлов применительно к безопасности различных режимов шифрования и ключевого расписания. В работах [12, 13] исследуется влияние S-блоков на лавинные эффекты, дифференциальные и линейные свойства блочных шифров. Работы [14 – 16] посвящены исследованию свойств нелинейных узлов замены в современных поточных шифрах в сравнении с алгоритмом «Strumok», предлагаемым в качестве нового стандарта поточного шифрования Украины [17]. Методы построения S-блоков исследуются многими авторами, например [18 – 20]. Однако наиболее распространенным и развитым остается математический аппарат криптографических булевых функций [21 – 28]. В частности, в [21] представлено новое рекурсивное построение булевой функции с максимальным алгебраическим иммунитетом; в [22, 23] рассмотрены генетические алгоритмы построения булевых функций с требуемыми криптографическими свойствами; в [24] исследуется метод имитации отжига; в [25, 26] исследуются эволюционные методы; работы [27, 28] посвящены эвристическим методам градиентного поиска.

Цель данной работы – продолжение исследований метода градиентного спуска, впервые предложенного в [28], оценка его вычислительной сложности в сравнении с наиболее близким аналогом из [27]. Для этого в разд. 1 вводятся необходимые термины и определения; в разд. 2 кратко излагаются исследуемые эвристические методы [27, 28] и приводятся расчетные данные необходимого числа операций для реализации градиентного спуска (табл. 1). В разд. 3 оцениваются свойства метода градиентного подъема по формированию высоко нелинейных корреляционно-иммунных криптографических булевых функций. В разд. 4 предлагается методика оценки эффективности эвристических методов и приводятся результаты сравнительных исследований. В частности, показано, что метод градиентного спуска из [28] за значительно меньшее число итераций (в десятки раз) позволяет формировать криптографические булевы функции с требуемыми показателями нелинейности и автокорреляции. В разд. 5 приводятся результаты исследований криптографических свойств формируемых булевых функций, проводится сравнение с наилучшими известными оценками. В заключение полученные результаты обобщаются, кратко формулируются направления дальнейших исследований.

1. Показатели стойкости криптографических булевых функций

Введем основные понятия и определения математического аппарата булевой алгебры, используемые при оценке эффективности нелинейных узлов замен симметричных шифров [1 – 17].

Булевой функцией f от n переменных является функция [1 – 17], осуществляющая отображение из поля $GF(2^n)$ всех двоичных векторов $x = (x_1, \ldots, x_n)$ длины n в поле $GF(2)$. Обыч-

но булевы функции представляются в алгебраической нормальной форме (АНФ) и рассматриваются как сумма произведений составляющих координат.

Алгебраическая степень deg(f) является степенью самого длинного слагаемого функции, представленной в алгебраической нормальной форме. Алгебраическая степень отражает стойкость к аналитическим атакам, призванным свести данную функцию к криптографически слабой (линейной).

Последовательностью функции f называется (1,-1)-последовательность, определенная как $((-1)^{f(\alpha_0)}, (-1)^{f(\alpha_1)}, \ldots, (-1)^{f(\alpha_{2^n-1})})$ [1 – 17].

Таблицей истинности функции f называется (0,1)-последовательность, определенная как $(f(\alpha_0), f(\alpha_1), f(\alpha_{2^n-1}))$ [1-17].

Последовательность функции *f* является *сбалансированной*, если ее (0,1)-последовательность ((1,-1)-последовательность) содержит одинаковое количество нулей и единиц (единиц и минус единиц). Функция *f* является сбалансированной, если сбалансирована ее последовательность [1 – 17].

Эквивалентное определение сбалансированности [1 – 17]: функция *f* над $GF(2^n)$ является сбалансированной, если ее выходные значения являются равновероятными:

$$| \{x \mid f(x) = 0 \} | = | \{x \mid f(x) = 1 \}| = 2^{n-1}.$$

Сбалансированность функции является показателем стойкости, отражающим слабость выходной последовательности к статистическим атакам.

Аффинной функцией f называется функция вида $f = a_1x_1 \oplus \ldots \oplus a_nx_n \oplus c$, где $a_j, c \in GF(2), j = 1, 2,..., n$. Функция *f* называется *линейной*, если $c = 0$ [1 – 17].

Весом Хэмминга вектора α ((0,1)-последовательности α), обозначаемым как $W(\alpha)$, является количество единиц в векторе (последовательности) [1 – 17].

Расстоянием Хэмминга d(f,g) между последовательностями двух функций *f* и *g* является количество позиций, в которых различны последовательности этих функций [1 – 17].

Нелинейность N_S преобразования – минимальное расстояние Хэмминга между выходной последовательностью *S* и всеми выходными последовательностями аффинных функций над некоторым полем [1 – 17]: $N_S = min\ \{d(S,\varphi)\}$, где φ – множество аффинных функций.

Нелинейность функции N_f – минимальное расстояние Хэмминга N_f между функцией *f* и всеми аффинными функциями над $GF(2^n)$ [1 – 17], где φ – множество аффинных функций.

Для произвольной функции *f* нелинейность N_f над $GF(2^n)$ может достигать [1 – 17]: $N_f \le 2^{n-1} - 2^{n/2-1}$.

Для сбалансированной функции *f* над $GF(2^n)$ ($n \ge 3$) нелинейность N_f может достигать [1 – 17]:

$$N_f \le \begin{cases} 2^{n-1} - 2^{n/2-1} - 2, n = 2k, \\ \left\lfloor\left\lfloor 2^{n-1} - 2^{n/2-1} \right\rfloor\right\rfloor, n = 2k+1, \end{cases}$$

где $\lfloor\lfloor x \rfloor\rfloor$ – максимальное четное целое, меньшее либо равное *x*.

Нелинейность функции является показателем, отражающим стойкость функций к корреляционным (линейным) атакам.

Функция *f* обладает *корреляционным иммунитетом* порядка *k*, если выходная последовательность функции $y \in Y$ статистически не зависит от любого подмножества из *k* входных координат [1 – 17]:

$$\forall \{x_1, \ldots, x_k\} \quad P(y \in Y / \{x_1, \ldots, x_k\} \in X) = P(y \in Y).$$

Эквивалентное определение корреляционного иммунитета в терминах преобразования Уолша [1 – 17]: функция *f* над полем $GF(2^n)$ имеет корреляционный иммунитет порядка *k*, *КИ*(*k*), если ее преобразование Уолша удовлетворяет равенству $F(\omega) = 0$ для всех $\omega \in V_n$ таких, что $1 \le W(\omega) \le k$: $\forall \omega \in V_n, F(\omega) = 0$, *КИ*(*f*) = *k*.

Преобразование Уолша $F(\omega)$ функции f над полем $GF(2^n)$ определяется как принимающая действительные значения функция [1 – 17]:

$$F(\omega) = 2^{-n} \sum_x (-1)^{f(x) \oplus \langle \omega, x \rangle},$$

где $\omega \in V_n$, $f(x)$, $\langle \omega, x \rangle \in N$ ($\langle \omega, x \rangle$ – скалярное произведение $w_1 x_1 \oplus \ldots \oplus w_n x_n$) .

Корреляционно-иммунная функция k-го порядка – функция, обладающая корреляционным иммунитетом порядка k. Сбалансированные корреляционно-иммунные функции называются *эластичными функциями*.

Функция f над полем $GF(2^n)$ удовлетворяет [1 – 17]:

- критерию распространения относительно вектора α, $KP(\alpha)$, если функция $f(\mathrm{x}) \oplus f(\mathrm{x} \oplus \alpha)$ является сбалансированной, $x \in V_n$, где $x = (x_1, x_2, \ldots, x_n)$:

$$P(f(\mathrm{x}) = f(\mathrm{x} \oplus \alpha)) = \frac{1}{2};$$

- критерию распространения степени k, $KP(k)$, если удовлетворяется критерий распространения относительно всех векторов $\alpha \in V_n$ при $1 \le W(\alpha) \le k$:

$$P(f(\mathrm{x}) = f(x \oplus \alpha)) = \frac{1}{2} \qquad , \forall \alpha : 1 \le W(\alpha) \le k;$$

- строгому лавинному критерию, *СЛК*, если f удовлетворяет критерию распространения степени 1:

$$P(f(\mathrm{x}) = f(\mathrm{x} \oplus \alpha)) = \frac{1}{2}, \ \forall \alpha : W(\alpha) = 1.$$

Степень корреляционного иммунитета/критерия распространения отражает стойкость функций к корреляционным атакам, призванным найти линейные свойства данной функции.

Функция f над $GF(2^n)$ называется бент-функцией [1 – 17], если

$$2^{-n/2} \sum_{x \in V_n} (-1)^{f(x) \oplus \langle \beta, x \rangle} = \pm 1.$$

для всех $\beta \in V_n$.

Последовательность бент-функции называется бент-последовательностью. Для бент-функций справедливы следующие утверждения [1 – 17]:

$\langle \xi, \ell \rangle = \pm 2^{n/2}$ для любой аффинной последовательности ℓ длины 2^n;

$f(x) \oplus f(x \oplus \alpha)$ сбалансирована $\forall \alpha \in V_n$, $W(\alpha) \ne 0$;

$f(x) \oplus \langle \alpha, x \rangle$ принимает значение единица $2^{n-1} \pm 2^{n/2-1}$ раз $\forall \alpha \in V_n$;

$f(x) \oplus h(x)$, где $h(x)$ – аффинная функция, также является бент-функцией.

Автокорреляционная функция $\hat{r}(s)$ для $s \in 0 \ldots 2^n - 1$ определена как

$$\hat{r}(s) = \sum_{x=0}^{2^n-1} \hat{f}(x) \hat{f}(x \oplus s).$$

Значение автокорреляции отражает стойкость функций к классу аналитических атак, призванным найти корреляцию между фрагментами функции.

Говорят, что функция f удовлетворяет характеристике распространения m, если

$$\left(1 \le |s| \le m\right) \Rightarrow |\hat{r}(s)| = 0.$$

Аналогично, автокорреляция $AC(f)$ функции f определяется как модуль наибольшего значения $\hat{r}(s)$:

$$AC(f) = \max_{s \ne 0} \left| \sum_{ч} \hat{f}(x) \hat{f}(x \oplus s) \right| = \max_{s \ne 0} |\hat{r}(s)|.$$

Автокорреляция $\hat{r}(s)$ обеспечивает утечку информационного потока со входа на выход функции.

2. Эвристические методы градиентного поиска

В данной работе исследуются эвристические методы градиентного поиска. В частности, метод градиентного подъема В. Миллана, Э. Кларка, Э. Доусона, 1997 г. [16] и разработанный на его основе метод градиентного спуска [17].

2.1. Эвристический метод градиентного подъема

Суть метода состоит в повышении нелинейности произвольной булевой функции путем комплементации некоторой позиции в таблице истинности исходной функции. Каждая позиция таблицы истинности соответствует уникальным входным данным. Метод позволяет создать полный список/перечень таких входных данных функции, что комплементация любой соответствующей данному входу выходной позиции в таблице истинности будет увеличивать нелинейность данной функции. Список/перечень таких позиций в таблице истинности обозначается как 1 – *Improvement Set* функции $f(x)$, или 1 – IS_f [16].

Определение 1 [16]. Пусть $g(x) = f(x) \oplus 1$ для $x = x_a$ и $g(x) = f(x)$ для всех остальных x. Если $N_g > N_f$, то $x_a \in 1 - IS_f$.

В [16] представлен быстрый систематический метод определения множества 1 – IS_f заданной булевой функции путем использования ее таблицы истинности и преобразований Уолша – Адамара. Для нахождения множества 1 – IS_f заданной булевой функции необходимо сначала определить значения коэффициентов преобразования Уолша – Адамара, которые соответствовали бы величинам, близким к абсолютному значению максимального коэффициента, WH_{max}.

Определение 2. Пусть $f(x)$ является булевой функцией с преобразованием Уолша – Адамара $F(w)$, где WH_{max} обозначает максимальное абсолютное значение $F(w)$. Тогда будут существовать одна или более линейных функций $L_w(x)$, имеющих минимальное расстояние до функции $f(x)$, и для данных w будет справедливо равенство $|F(w)| = WH_{max}$.

Определяется следующее множество:

$$W_1^+ = \{ w: F(w) = WH_{max} \} \text{ и}$$

$$W_1^- = \{ w: F(w) = - WH_{max} \}.$$

Также определяются множества w, для которых значения *WHT* приближены к максимуму:

$$W_2^+ = \{ w: F(w) = WH_{max} - 2\},$$

$$W_2^- = \{ w: F(w) = - (WH_{max} - 2)\},$$

$$W_3^+ = \{ w: F(w) = WH_{max} - 4\} \text{ и}$$

$$W_3^- = \{ w: F(w) = - (WH_{max} -4)\}.$$

Когда таблица истинности изменяется ровно в одном месте, все *WHT* значения изменяются на +2 или -2. Из этого следует, что для увеличения нелинейности все *WHT* значения в множестве W_1^+ должны быть изменены на -2, все *WHT* значения в множестве W_1^- должны быть изменены на 2, а также все *WHT* значения в множестве W_2^+ должны быть изменены на -2, все *WHT* значения в множестве W_2^- должны быть изменены на 2. Если первые два условия являются очевидными, то следующие два условия требуются для того, чтобы все другие значения $|F(w)|$ оставались меньшими, чем WH_{max}. Данные условия могут быть представлены в виде простых тестов.

Теорема 1 [16]. Пусть дана некоторая булева функция $f(x)$ с *WHT* $F(w)$ и определены множества $W^+ = W_1^+ \cup W_2^+$ и $W^- = W_1^- \cup W_2^-$. Тогда для некоторого входа x существует элемент из *Improvement Set* и выполняются следующие два условия:

$$f(x) = L_w(x) \text{ для всех } w \in W^+,$$

и $f(x) \neq L_w(x)$ для всех $w \in W$.

Если функция $f(x)$ не сбалансирована, понижение несбалансированности может быть достигнуто использованием дополнительного ограничения:

если $F(0) > 0$, $f(x) = 0$, иначе $f(x) = 0$.

Критерием градиентного поиска является максимизация расстояния по Хеммингу между формируемой последовательностью и последовательностями линейных функций. После обновления алгебраической формы булевой функции производятся аналогичные операции: выполняется преобразование Уолша – Адамара *WHT* и находится максимальные значения коэффициентов преобразования; формируется множество *Improvement Set*; находятся элементы последовательности функции, совпадающие с элементами последовательности ближайшей линейной формы; инвертирование совпавших элементов и повышение нелинейности функции, посредством «отдаления» от ближайшей линейной функции. Далее выполняются очередные итерации, аналогичные рассмотренным выше.

Проведенные исследования показали, что рассмотренный метод градиентного подъема вычислительно затратен и, при большом числе аргументов булевой функции, требует выполнения значительного числа повторяющихся итераций. Для снижения вычислительной сложности в [17] предложен метод градиентного спуска с бент-последовательностями в качестве входных данных.

2.2. Эвристический метод градиентного спуска

Данный метод основан на комплементации позиций бент-последовательностей для градиентного поиска сбалансированных булевых функций по критерию максимизации расстояния Хемминга между формируемыми последовательностями и последовательностями всех линейных функций. Это позволяет снизить вычислительные затраты на поиск булевых функций с требуемыми криптографическими свойствами.

Основной идеей метода градиентного спуска является эффективное понижение нелинейности заданных бент-последовательностей при каждой из 2n/2–1 обязательных комплементаций. В табл. 1 представлены расчетные данные для векторных пространств V4 – V12. В столбце 2 указана нелинейность (значение преобразования Уолша) бент-последовательностей, рассматриваемых как входные данные, в столбце 3 указана максимально достижимая нелинейность функций (максимальное значение преобразования Уолша), которые мы хотели бы получить в качестве выходных данных, и в столбце 4 указано количество бит, которое необходимо изменить в бент-последовательностях для получения желаемого результата.

Таблица 1

Расчетные значения для векторных пространств $V_4 – V_{12}$

	Максимально достижимые показатели для бент-функций		Максимально достижимые показатели для сбалансированных функций / Наилучший известный результат		Необходимо изменить позиций в бент-последовательности
	N_f	$F(w)$	N_f	$F(w)$	
V_4	6	4	4/4	8/8	2 позиции
V_6	28	8	26/26	12/12	4 позиции
V_8	120	16	118/116	20/24	8 позиции
V_{10}	496	32	494/492	36/40	16 позиции
V_{12}	2016	64	2014/2010	68/76	32 позиции

На рис. 1 представлены возможные потери нелинейности при комплементации необходимого числа позиций бент-последовательности.

Для достижения заданной верхней границы нелинейности необходимо из общего числа позиций x таблицы истинности, подлежащих комплементации, определить то число позиций y, изменение которых повлечет изменение WH на +2, и то число позиций z, изменение которых повлечет изменение WH на -2, $x = y + z$. В табл. 1 представлены расчетные данные, отображающие необходимое число требуемых комплементаций бент-последовательности для заданного векторного пространства в соответствии с теоремой 2.1 из [17].

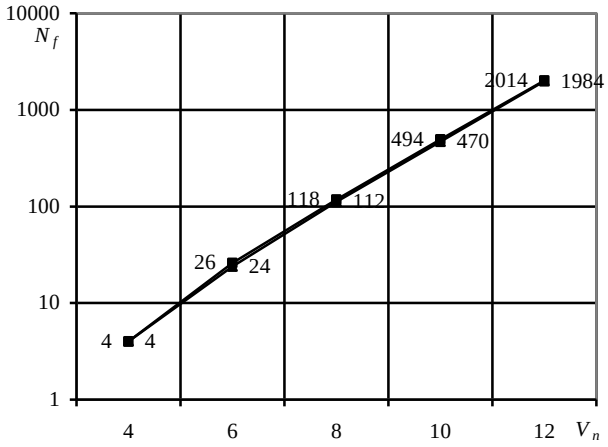

Рис. 1. Возможные потери нелинейности при комплементации

Таблица 2

Расчетные данные необходимого числа комплементаций бент-последовательности

	Необходимо изменить позиций в бент-последовательности, *NeedSteps*	Необходимо изменить значение нелинейности		Требуется для этого изменений, n^- и n^+
		N_f, с _ на _	$F(w)$, с _ на _	
V_4	2	6→4	4→8	$n^- = 2$ (изменения с $F(w) = +2$)
V_6	4	28→26	8→12	$n^- = 3$ (изменения с $F(w) = +2$) $n^+ = 1$ (изменения с $F(w) = -2$)
V_8	8	120→116	16→24	$n^- = 6$ (изменения с $F(w)$=+2) $n^+ = 2$ (изм-я с $F(w)$=-2)
V_{10}	16	496→492	32→40	$n^- = 10$ (изменения с $F(w)$=+2) $n^+ = 6$ (изменения с $F(w)$=-2)
V_{12}	32	2016→2010	64→76	$n^- = 19$ (изменения с $F(w)$=+2) $n^+ = 13$ (изменения с $F(w)$=-2)

После расчета необходимого числа комплементаций бент-последовательности на первом шаге эвристического поиска выполняется преобразование Уолша – Адамара WH и определяется максимальное расстояние по Хеммингу к одной или нескольким последовательностям линейных функций $L_i(x)$. Эта операция соответствует выбору нулевого значения коэффициентов преобразования Уолша – Адамара WH, после чего формируется множество линейных функций, составляющих *Improvement Set*. Далее производится инвертирование элементов последовательности бент-функции, совпадающих с элементами последовательностей линейных функций из множества *Improvement Set*. В результате несбалансированность функции снижается, но снижается также и нелинейность, т.е. последовательность функции не является уже максимально отдаленной от последовательностей линейных функций $L_i(x)$. На следующей итерации все операции повторяются. Таким образом, в качестве критерия градиентного поиска криптографических функций предлагаемым методом является макимизация минимального расстояния по Хеммингу формируемой последовательности и последовательностей линейных функций.

В целом предлагаемый метод структурно состоит из трех основных этапов.

На первом этапе используются процедуры градиентного спуска, позволяющие получить высоко нелинейную последовательность.

На втором этапе используется процедуры восстановления алгебраической нормальной формы функции по выходной последовательности.

На третьем этапе, в зависимости от среды практического приложения, используется процедура модификации алгебраической нормальной формы функции *f(x)*. Это позволяет при сохранении основных показателей стойкости (сбалансированности и нелинейности) путем применения аффинных преобразований улучшить либо динамические свойства нелинейного преобразования, либо корреляционные характеристики.

Таким образом, разработанный метод позволяет формировать сбалансированные криптографические функции с высокими показателями нелинейности. При этом, как показано на рис. 1, значения нелинейности лежат в узком диапазоне значений, который зависит от размерности векторного пространства.

Следует отметить, что для современных поточных шифров важным показателем эффективности является также корреляционная иммунность, характеризующая устойчивость схемы шифрования к корреляционным атакам. Проведем оценку нелинейности и корреляционного иммунитета булевых функций, которые могут быть синтезированы разработанным методом.

3. Оценка нелинейности и корреляционного иммунитета формируемых функций

Для криптографических булевых функций известна взаимосвязь между достижимой степенью корреляционного иммунитета m и ее нелинейностью N_f [30]:

$$N_f = 2^{n-1} - 2^{m+1}, \quad (1)$$

справедливая для

$$m \geq n/2 - 2. \quad (2)$$

Как видно из (1), повышение степени корреляционного иммунитета m ведет к понижению нелинейности, и наоборот. Поэтому разработчикам средств криптографической защиты в зависимости от условий практического использования приходится находить компромисс между требуемой нелинейностью и желаемой степенью корреляционного иммунитета. Достоинство разработанного метода состоит в возможности строить функции с различными значениями криптографических показателей.

Так, например, в табл. 3 на основе (1) и (2) представлена достижимая степень корреляционного иммунитета $CI_{\max}(k)$ с указанием соответствующей нелинейности $N_{f\,\min}$. Фактически приведенные в таблице данные соответствуют нижней границе нелинейности, гарантированно получаемой при использовании разработанного метода. В табл. 4 приведена достижимая степень корреляционного иммунитета $CI_{\max}(k)$ с указанием максимально возможной нелинейности для сбалансированных функций $N_{f\,\max}$, т.е. здесь приведена верхняя граница нелинейности функций при использовании разработанного метода. На рис. 2 для наглядности табличные данные изображены в виде диаграммы.

Таблица 3

Нижняя граница нелинейности
при заданном корреляционном иммунитете

	V_4	V_6	V_8	V_{10}	V_{12}
$CI_{\max}(k)$	1	2	3	4	5
$N_{f\,\min}$	4	24	112	480	1984

Таблица 4

Верхняя граница нелинейности при заданном корреляционном иммунитете

	V_4	V_6	V_8	V_{10}	V_{12}
$CI_{\max}(k)$	1	1	2	3	4
$N_{f\max}$	4	26	116	492	2010

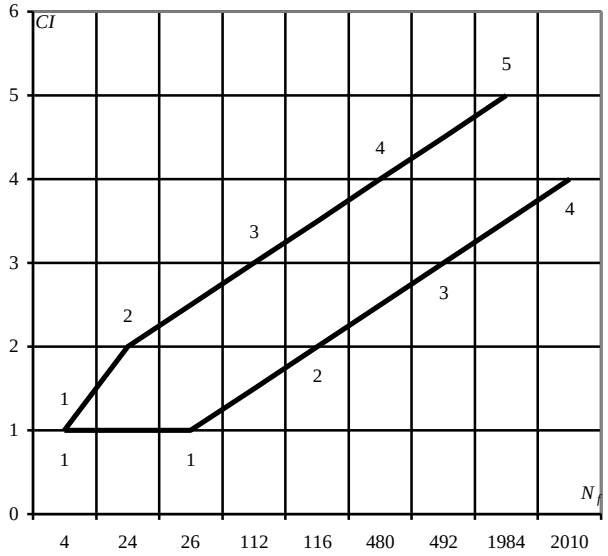

Рис. 2. Граничные показатели корреляционного иммунитета

Как показывает анализ приведенных данных, применение разработанного метода позволяет формировать булевы функции, которые, помимо высоких значений нелинейности, потенциально могут быть корреляционно-иммунными функциями. При их применении в поточных шифрах будет обеспечена высокая устойчивость к различным криптографическим атакам. Так, например, применение разработанного метода над пространством V_8 позволяет формировать функции с показателем нелинейности $N_{f\min} = 112$ и степенью корреляционного иммунитета $CI_{\max}(3)$, что является лучшим известным на сегодняшний день результатом.

Следует отметить, что вероятностный поиск эвристическими методами описывается некоторым случайным процессом, конкретная реализация которого суть случайные величины – значения показателей стойкости найденной функции (см. п. 1). Соответствующие вероятности наступления искомых случайных событий указывают на среднее число попыток до успеха – построения криптографической булевой функции с требуемыми свойствами. Таким образом, для оценки вычислительной эффективности эвристических методов, т.е. оценки соответствия полученного результата требуемому, необходимо провести оценку распределения вероятностей формирования булевых функций с различными криптографическими показателями.

4. Методика оценки эффективности эвристических методов и результаты исследований

Предлагаемая методика использует в качестве показателя вычислительной эффективности среднее число попыток, которое потребуется выполнить с использованием эвристического метода, для формирования криптографической функции с требуемыми показателями стойкости.

В соответствии с основными положениями теории вероятности и математической статистики неизвестную функцию распределения рассматриваемой случайной величины определяют по результатам наблюдений, по выборке [18]. Выборкой объема L для случайной величины A называется последовательность $X_1, X_2, \ldots, X_L$ из L независимых наблюдений этой

величины, т.е. совокупность значений, принятых L независимыми случайными величинами $A_1, A_2, \ldots, A_L$, имеющими тот же закон распределения $F_A(x)$, что и рассматриваемая величина A. В этом случае говорят, что выборка $X_1, X_2, \ldots, X_L$ взята из генеральной совокупности величины A, а под законом распределения генеральной совокупности понимают закон распределения случайной величины A. Значения $X_1, X_2, \ldots, X_L$ называют выборочными значениями [18].

Введем следующие обозначения: SI_i – случайная величина, значения которой представляют собой исходы эвристического поиска – численное выражение i-го показателя стойкости криптографической булевой функции; $X_1, X_2, \ldots, X_L$ – выборка объема L случайной величины SI_i; $F_{SI_i}(x)$ – функция распределения случайной величины SI_i.

Оценим значения теоретических функций распределения $F_{SI_i}(x)$, являющихся вероятностями событий $\{SI_i < x\}$, с помощью частот этих событий по выборке объема L. Обозначим через v_x количество выборочных значений, меньших x. Тогда $\frac{v_x}{L}$ частоты попадания выборочных значений левее точки x в данной выборке, т.е. частоты событий $\{SI_i < x\}$. Эти частоты являются функциями от x и являются, соответственно, эмпирическими функциями распределения $F^*_{SI_i}(x)$ случайных величин SI_i, полученными по данной выборке: $F^*_{SI_i}(x) = \frac{v_x}{L}$. Частота события в L независимых опытах является оценкой для вероятности этого события, т.е.

$$F_{SI_i}(x) \approx F^*_{SI_i}(x) = \frac{v_x}{L}.$$

Используя функцию распределения $F_{SI_i}(x)$, введем показатель вычислительной эффективности эвристических методов как среднее число K_{cp} попыток вероятностного формирования булевой функции с требуемыми свойствами:

$$K_{cp} = \frac{1}{F_{SI_i}(x)} \approx \frac{1}{F^*_{SI_i}(x)}.$$

Если принять предположение о статистической независимости m случайных величин SI_i, $i = 1,..,m$, тогда вероятность формирования криптографической функции с показателями $SI_i < x, i = 1,..,m$ будет определяться вероятностью совместного события, записанной через произведение вероятностей независимых событий: $\prod_{i=1}^{m} F_{SI_i}(x)$.

Среднее число попыток вероятностного формирования криптографической функции с $SI_i < x, i = 1,..,m$ вычислим по выражению

$$K_{cp} = \frac{1}{\prod_{i=1}^{m} F_{SI_i}(x)} \approx \frac{1}{\prod_{i=1}^{m} F^*_{SI_i}(x)}.$$

Наибольший интерес в криптографических целях представляют два основных показателя: нелинейность N_f и автокорреляция AC [1 – 17], причем необходимо максимизировать нелинейность и минимизировать автокорреляцию. Для оценки вычислительной эффективности по этим двум показателям стойкости последнее выражение перепишем в виде

$$K_{cp} = \frac{1}{\left(1 - F_{N_f}(x)\right) \cdot F_{AC}(x)} \approx \frac{1}{\left(1 - F^*_{N_f}(x)\right) \cdot F^*_{AC}(x)},$$

где $F_{N_f}(x)$ и $F^*_{N_f}(x)$ – теоретическое и эмпирическое значение вероятностей наступления события $\{N_f \le x\}$; $F_{AC}(x)$ и $F^*_{AC}(x)$ – теоретическое и эмпирическое значение вероятностей наступления события $\{AC \le x\}$;

Используя показатель K_{cp}, проведем сравнительные исследования вычислительной эффективности эвристических методов вероятностного формирования криптографических

булевых функций. В качестве объекта исследования будут выступать метод случайной генерации [1 – 17], метод градиентного подъема и предложенный в [16] эвристический метод градиентного спуска [17].

На рис. 3 представлены гистограммы частот событий $\{N_f = x\}$ для сбалансированных булевых функций, построенных над V_8, объем выборки L =10000. Как видно из приведенных данных, эвристический метод градиентного спуска (IKK) позволяет формировать булевы функции с показателями нелинейности $N_f \geq 114$ с вероятностью 1, $N_f \geq 116$ с вероятностью 0.5. Следующий за ним по вычислительной эффективности метод градиентного подъема (MSD) позволяет формировать криптографические функции с показателями нелинейности $N_f \geq 112$ с вероятностью 1, $N_f \geq 114$ с вероятностью 0.5 и $N_f \geq 116$ с вероятностью 0.1. Метод же случайной генерации (RG) является вообще малоэффективным, наиболее вероятное значение нелинейности находится в диапазоне 80 – 104.

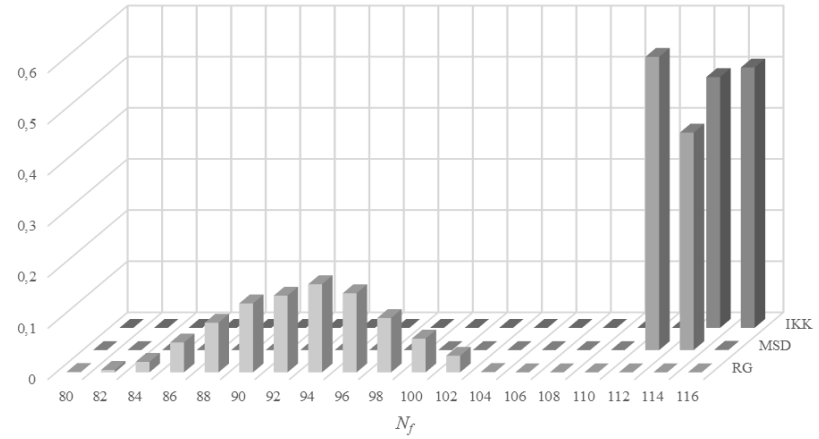

Рис. 3. Гистограммы частот событий $\{N_f = x\}$, объем выборки L = 10000

На рис. 4 представлены гистограммы частот событий $\{AC = x\}$ для сбалансированных булевых функций, построенных над V_8, объем выборки L =10000.

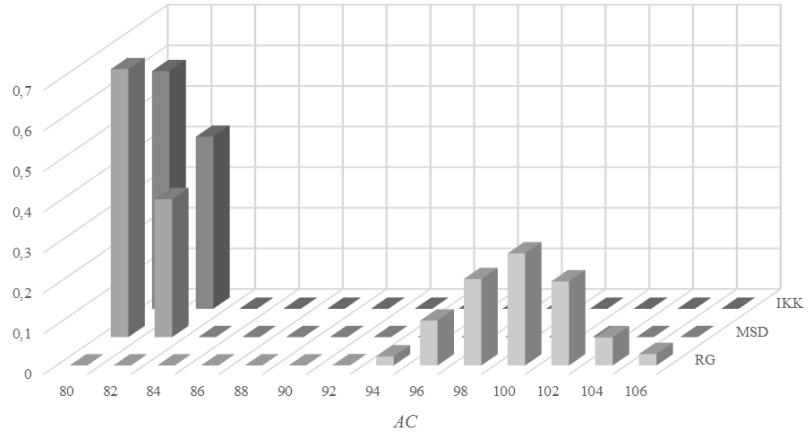

Рис. 4. Гистограммы частот событий $\{AC = x\}$, объем выборки L = 10000

Как показывает анализ, эвристический метод градиентного спуска не уступает ближайшему аналогу – методу градиентного поиска. Он позволяет формировать булевы функции с низким показателем автокорреляции.

На рис. 5 представлены зависимости K_{cp} для: метода случайной генерации с AC = 80 (RG, AC=80); метода случайной генерации с AC = 120 (RG, AC=120); метода градиентного подъема с AC = 24 (MCD, AC=24); метода градиентного подъема с AC = 32 (MCD, AC=32); метода градиентного спуска с AC = 24 (IKK, AC=24); метода градиентного спуска с AC = 32 (IKK, AC=24).

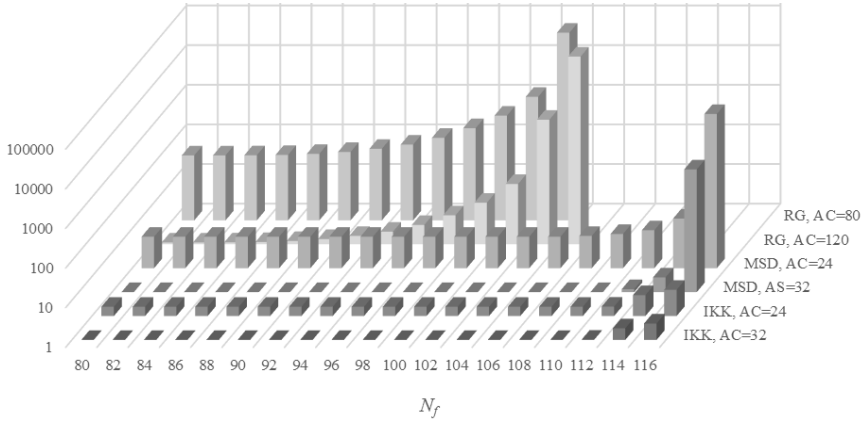

Рис. 5. Зависимости среднего числа $\kappa_{ср}$

Анализ зависимостей, приведенных на рис. 4, показывает, что метод градиентного спуска позволяет формировать булевы функции с высокими криптографическими показателями (нелинейностью и автокорреляцией) за меньшее число попыток (в среднем). Так, например, формирование криптографической функции с AC = 24 и N=116 для метода случайной генерации вычислительно недостижимо по причине чрезвычайно высокого среднего числа попыток. Для тех же параметров метод градиентного подъема потребует в среднем около 8000 попыток. Метод градиентного спуска при тех же показателях потребует в среднем четыре попытки, т.е. среднее число попыток снизилось в 2000 раз. При требованиях к криптографическим свойствам AC = 24 и N=114 метод градиентного подъема потребует в среднем около 15 попыток, а метод градиентного спуска – около 3.

5. Криптографические свойства формируемых булевых функций

Проведем сравнительные исследования свойств криптографических булевых функций с наилучшими известными аналогами: генетическим алгоритмом [31], алгоритмами NLT и ACT [32], которые относятся к классу эвристических методов.

В табл. 5 представлены результаты сравнительной оценки нелинейности функций, полученные при использовании разработанного метода градиентного спуска, метода-прототипа (эвристического метода градиентного подъема) и наилучших известных эвристических методов (все данные, за исключением последней строки, взяты из [31]).

Приведенные данные свидетельствуют, что среди эвристических методов разработанный метод позволяет достигать наивысшей нелинейности. Высокая нелинейность свидетельствует о высокой степени замешивания данных, что определяет стойкость криптопреобразований. Нам впервые удалось построить функции с наивысшей известной нелинейностью среди эвристических методов: $N_f = 488$ для V_{10} и $N_f = 2002$ для V_{12}.

Таблица 5

Сравнительная оценка нелинейности функций

	V_6	V_8	V_{10}	V_{12}
Теоретически достижимая нелинейность	26	118	494	2014
Метод случайной генерации [31]	-	112	472	1954
Hill Climbing Method [27]	-	114	476	1960
Genetic Algorithm [31]	26	116	484	1976
NLT [32]	26	116	486	1992
ACT [32]	26	116	484	1986
Разработанный метод [28]	26	116	488	2002

В табл. 6 приведены сравнительные характеристики наилучших известных методов, позволяющих строить функции с низкими значениями автокорреляции [31]. Как видно из приведенной таблицы, разработанный метод позволяет строить функции с низкими значе-

ниями автокорреляции. Над V_8 методы *NLT* и *ACT* позволяют строить функции с AC=16, однако при этом нелинейность равна 112. Разработанный метод позволяет строить функции с нелинейностью 116. Над всеми остальными векторными пространствами полученные значения сопоставимы с результатами для других методов.

Таблица 6

Сравнительная оценка автокорреляции функций

	V_6	V_8	V_{10}	V_{12}
Zhang Zheng [33, 34]	16	24	48	96
Maitra [35, 36]	16	24	40	80
NLT [32]	16	16	64	144
ACT [32]	16	16	56	128
Разработанный метод [28]	16	24	40	72

На рис. 6 – 10 представлены спектральные свойства булевых функций, построенных различными способами. В скобках приведены показатели: (n, $deg(f)$, N_f, AC).

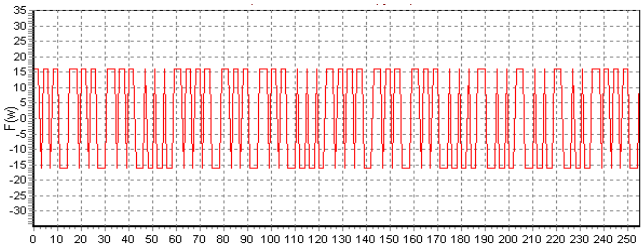

Рис. 6. Бент-функция [31 – 40]: (8, 4, 120, 0)

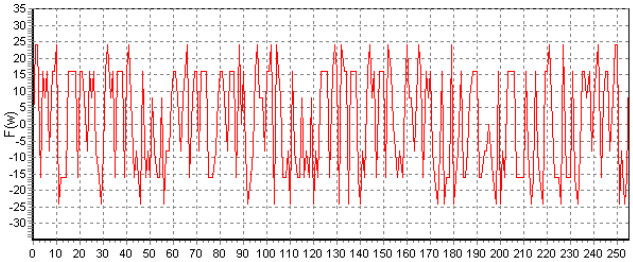

Рис. 7. Разработанный метод [28]: (8, 7, 116, 24)

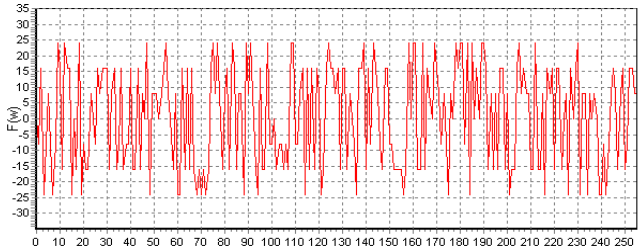

Рис. 8. Hill Climbing Method [28]: (8, 6, 116, 24)

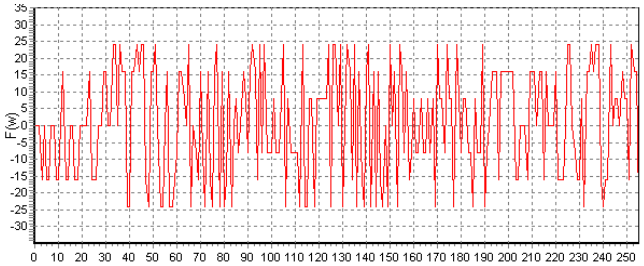

Рис. 9. Метод Maitra – Pasalic [37]: (8, 6, 116, 80)

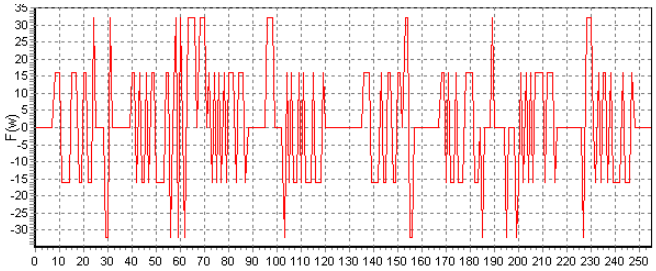

Рис. 10. Метод Seberry-Zhang [38 – 40]: (8, 4, 112, 128)

Приведенные данные показывают, что криптографические булевы функции, построенные в соответствии с разработанным методом [28], имеют максимально достижимую алгебраическую степень, высокую нелинейность низкую автокорреляцию. По большинству показателей стойкости сформированные функции не уступают известным методам.

Выводы

Проведенные исследования вычислительной эффективности эвристических методов показали, что методы градиентного поиска позволяют за приемлемое число итераций формировать криптографические булевы функции с высокими показателями нелинейности и низкой автокорреляцией. Метод градиентного спуска, предложенный в [17], оказался эффективнее Hill Climbing Method из [16]. В частности, результаты экспериментальных исследований показывают, что метод градиентного спуска из [28] требует в десятки раз меньшее число итераций, т.е. он значительно эффективнее в вычислительном аспекте. Предложенная методика оценки вычислительной эффективности эвристических методов может быть использована и для других методов, в том числе использующих расширенный набор показателей стойкости.

Сравнительные исследования криптографических свойств булевых функций показали, что формируемые предложенным вычислительным методом функции обладают высокими показателями: показатель нелинейности приближается к верхней теоретической границе; показатель автокорреляции является одним из самых низких по сравнению с другими методами синтеза; при равных показателях нелинейности формируемые функции имеют максимально достижимую алгебраическую степень; все известные методы синтеза уступают по спектральным характеристикам функций. Таким образом, на основе проведенных исследований можно сделать вывод о том, что функции, построенные в соответствии с разработанным методом [28], имеют высокие показатели стойкости и превосходят по данным показателям известные функции.

Перспективным направлением дальнейших исследований является разработка вероятностной модели синтеза нелинейных узлов замен с высокими криптографическими свойствами, проведение экспериментальных исследований и обоснование практических рекомендаций по внедрению полученных результатов.

Список литературы:

1. Information technology. Security techniques. Encryption algorithms. Part 3: Block ciphers. ISO/IEC 18033-3: 2010, 2010.

2. Advanced Encryption Standard. Federal Information Processing Standards Publications FIPS-197, 2001.

Information technologies. Cryptographic Data Security. Symmetric block transformation algorithm. National Standard of Ukraine DSTU 7624:2014, 2015 (in Ukr.).

3. Information technology. Cryptography protection of information. Block ciphers. National Standard of Russian Federation GOST R 34.12-2015, 2015 (in Rus.).

4. Information technology and security. Information security. Cryptography encryption and integrity control algorithms. State Standard of the Republic of Belarus STB 34.101.31-2011, 2011.

5. O.O. Kuznetsov, Yu.I. Gorbenko, I.M. Bilozertsev, A.V. Andrushkevych, O.P. Narizhnyi. Algebraic Immunity of Non-linear Blocks of Symmetric Ciphers // Telecommunications and Radio Engineering. – 2018. – Vol. 77, Issue 4. – P. 309-325.

6. B. N. Tran, T. D. Nguyen and T. D. Tran. A New S-Box Structure to Increase Complexity of Algebraic Expression for Block Cipher Cryptosystems // International Conference on Computer Technology and Development, Kota Kinabalu, 2009. – P. 212-216.

7. A. Kuznetsov, R. Serhiienko, D. Prokopovych-Tkachenko and Y. Tarasenko. Evaluation of Algebraic Immunity of modern block ciphers // IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kyiv, Ukraine, 2018. – P. 288-293.

8. M. McLoone and J. V. McCanny. High-performance FPGA implementation of DES using a novel method for implementing the key schedule // IEE Proceedings – Circuits, Devices and Systems, vol. 150, no. 5, pp. 373, 6 Oct. 2003.

9. A. Kuznetsov, I. Kolovanova and T. Kuznetsova. Periodic characteristics of output feedback encryption mode // 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kharkov, 2017. – P. 193-198.

10. S. Sulaiman, Z. Muda and J. Juremi. The new approach of Rijndael key schedule // Proceedings Title: 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), Kuala Lumpur, 2012, pp. 23-27.

11. O. Kuznetsov, Y. Gorbenko and I. Kolovanova. Combinatorial properties of block symmetric ciphers key schedule // Third International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T), Kharkiv, 2016, pp. 55-58.

12. F. H. Nejad, S. Sabah and A. J. Jam. Analysis of avalanche effect on advance encryption standard by using dynamic S-Box depends on rounds keys // International Conference on Computational Science and Technology (ICCST), Kota Kinabalu, 2014, pp. 1-5.

13. H. Liu and C. Jin. Lower Bounds of Differential and Linear Active S-boxes for 3D-like Structure // The Computer Journal, vol. 58, no. 4, pp. 904-921, April 2015.

14. A. Kuznetsov, Y. Gorbenko, A. Andrushkevych and I. Belozersev. Analysis of block symmetric algorithms from international standard of lightweight cryptography ISO/IEC 29192-2 // 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kharkov, 2017, pp. 203-206.

15. I. Gorbenko, A. Kuznetsov, M. Lutsenko and D. Ivanenko. The research of modern stream ciphers // 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kharkov, 2017, pp. 207-210.

16. A. Kuznetsov, V. Frolenko, E. Eremin and O. Zavgorodnia. Research of cross-platform stream symmetric ciphers implementation // IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kyiv, Ukraine, 2018, pp. 300-305.

17. I. Gorbenko, O. Kuznetsov, Y. Gorbenko, A. Alekseychuk and V. Tymchenko, "Strumok keystream generator," 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kyiv, Ukraine, 2018, pp. 294-299.

18. V. Gopi and E. Logashanmugam. Design and analysis of nonlinear AES S-box and mix-column transformation with the pipelined architecture // International Conference on Current Trends in Engineering and Technology (ICCTET), Coimbatore, 2013, pp. 235-238.

19. H. Wang, H. Zheng, B. Hu and H. Tang. Improved Lightweight Encryption Algorithm Based on Optimized S-Box // International Conference on Computational and Information Sciences, Shiyang, 2013, pp. 734-737.

20. I. Das, S. Nath, S. Roy and S. Mondal. Random S-Box generation in AES by changing irreducible polynomial // International Conference on Communications, Devices and Intelligent Systems (CODIS), Kolkata, 2012, pp. 556-559.

21. Y. Chen, W. Tian and Y. Zhang. Construction for Balanced Boolean Function with Maximum Algebraic Immunity // 7th International Conference on Advanced Software Engineering and Its Applications, Haikou, 2014, pp. 32-34.

22. C. E, S. Liang and T. Zhang. Construction Method of Boolean Functions Based on Genetic Algorithm // 7th International Conference on Wireless Communications, Networking and Mobile Computing, Wuhan, 2011, pp. 1-4.

23. R. Asthana, N. Verma and R. Ratan. Generation of Boolean functions using Genetic Algorithm for cryptographic applications // IEEE International Advance Computing Conference (IACC), Gurgaon, 2014, pp. 1361-1366.

24. Bharti and D. K. Sharma. Searching boolean function using simulated annealing and hill climbing optimization techniques // International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), Ramanathapuram, 2016, pp. 62-64.

25. W. Millan, J. Fuller and E. Dawson. New concepts in evolutionary search for Boolean functions in cryptology // Evolutionary Computation, 2003. CEC '03. The 2003 Congress on, 2003, pp. 2157-2164 Vol.3.

26. S. Picek, C. Carlet, S. Guilley, J. F. Miller and D. Jakobovic. Evolutionary Algorithms for Boolean Functions in Diverse Domains of Cryptography // Evolutionary Computation, vol. 24, no. 4, pp. 667-694, Dec. 2016.

27. W. Millan, A. Clark, E. Dawson. Smart Hill Climbing Finds Better Boolean Functions // Proceedings of the Workshop on Selected Areas on Cryptography SAC 97, Springer-Verlag, pp. 50-63, 1997.

28. Y. Izbenko, V. Kovtun and A. Kuznetsov. The Design of Boolean Functions by Modified Hill Climbing Method // Sixth International Conference on Information Technology: New Generations, Las Vegas, NV, 2009, pp. 356-361.

29. A statistical test suite for random and pseudorandom number generators for cryptographic applications. NIST Special Publication 800-22, 2001.

30. E. Pasalic and T. Johansson. Further results on the relation between nonlinearity and resiliency of Boolean functions // Proc. IMA Conf. Cryptography and Coding (Lecture Notes in Computer Science). New York: Springer-Verlag, 1999, vol. 1746, pp. 35–45.

31. Clark J., Jacob S., Stepney S., Maitra, Millan W. Evolving Boolean Functions Satisfying Multiple Criteria // Proceedings of INDOCRYPT'02. LNCS Vol. 2551, Springer (2002) 246-259.

32. Millan W., Clark A., Dawson E. An Effective Genetic Algorithm for Finding Highly Non-linear Boolean Functions // Proceedings of the First International Conference on Information and Communications Security. LNCS Vol. 1334. Springer-Verlag, Berlin Heidelberg New York (1997) 149-158.

33. Y. Zheng and X. M. Zhang. Improved upper bound on the nonlinearity of high order correlation immune functions // Selected Areas in Cryptography-SAC 2000, Lecture Notes in Computer Science, Volume 2012, pages 264–274. Springer Verlag, 2000.

34. X-M. Zhang and Y. Zheng. GAC-the criterion for global avalanche characteristics of cryptographic functions // Journal of Universal Computer Science, 1(5):316–333, 1995.

35. S. Maitra. Highly nonlinear balanced Boolean functions with very good autocorrelation property // Workshop on Coding and Cryptography-WCC 2001, Paris, January 8–12, 2001. Electronic Notes in Discrete Mathematics, Volume 6, Elsevier Science, 2001.

36. S. Maitra. Autocorrelation properties of correlation immune Boolean functions // INDOCRYPT 2001, Lecture Notes in Computer Science Volume 2247, pages 242–253. Springer Verlag, December 2001.

37. S. Maitra and E. Pasalic. Further constructions of resilient Boolean functions with very high nonlinearity // IEEE Transactions on Information Theory, 48(7): 1825–1834, July 2002.

38. J. Seberry, X.-M. Zhang and Y.Zheng. Nonlinearity and Propagation Characteristics of Balanced Boolean Functions // Information and Computation, Vol. 119, No 1, pp. 1 – 13, 1995.

39. J.Seberry, X.M.Zhang, Y.Zheng. On Constractions and Nonlinearity of Correlation Immune Functions // Advances in Cryptology – EUROCRYPT'93, vol.765, Lecture Notes in Computer Science, Springer-Verlag, pp.181-199,1994.

40. J. Seberry and X. Zhang. Hadamar Matrices, Bent Functions and Cryptography // J.H.Dinitz and D.R. Stinson, editors, Contemporary Design Theory: A Collection of Surveys,chapter 11, pages 431-559, John Wiley and Sons, Inc, 1995.

Харьковский национальный
университет имени В.Н.Каразина;
Харьковский национальный университет
Воздушных Сил имени Ивана Кожедуба;
Университет таможенного дела и финансов, Днепр

Поступила в редколлегию 04.11.2018