ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

DOI 10.17223/20710410/13/1

УДК 631.391:519.2

БЫСТРЫЙ АЛГОРИТМ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ МАКСИМАЛЬНОЙ НЕСБАЛАНСИРОВАННОСТИ БИЛИНЕЙНЫХ АППРОКСИМАЦИЙ БУЛЕВЫХ ОТОБРАЖЕНИЙ

А. Н. Алексейчук, А. С. Шевцов

Институт специальной связи и защиты информации Национального технического университета Украины «Киевский политехнический институт», г. Киев, Украина

E-mail: alex-crypto@mail.ru, ashef@mail.ru

Предложен вероятностный алгоритм, позволяющий оценивать сверху максимальную несбалансированность (в заданном классе) билинейных аппросимаций булевых отображений n переменных за время, линейно зависящее от n.

Ключевые слова: блочный шифр, билинейный криптоанализ, булево отображение, билинейная аппроксимация, вероятностный алгоритм.

Введение

При исследовании стойкости блочных шифров относительно билинейного метода криптоанализа требуется вычислять или оценивать максимальные значения несбалансированности билинейных аппроксимаций булевых отображений, реализуемых узлами замены, раундовой функцией или шифром в целом [1-3]. Естественные алгоритмы решения этой задачи, основанные на «прямом» вычислении искомых параметров, имеют экспоненциальную трудоемкость и становятся практически неприменимыми уже при умеренных значениях числа переменных данного отображения.

Начиная с публикации О. Гольдрайха и Л. Левина [4], известен ряд вероятностных алгоритмов [5-10], позволяющих формировать список «высоковероятных» линейных аппроксимаций произвольной булевой функции n переменных за полиномиальное от n время. Применение таких алгоритмов к линейным комбинациям координатных функций булева отображения позволяет существенно уменьшить сложность нахождения его наиболее вероятных линейных аппроксимаций (за счет некоторого снижения достоверности результата, что обусловлено вероятностным характером применяемых алгоритмов). Отметим, в частности, работу [10], где с использованим одного из таких алгоритмов получено около 80 (близких по качеству к наилучшим известным) линейных аппроксимаций шифра DES с 8 раундами шифрования.

Следует подчеркнуть, что алгоритмы, изложенные в [4–10], предназначены именно для построения линейных аппроксимаций, несбалансированности которых ограничены снизу определенным значением. Вместе с тем в ряде задач криптографии, например при обосновании стойкости блочных шифров или их элементов относительно линейных атак, требуется находить лишь нетривиальные верхние оценки максимальной несбалансированности линейных аппроксимаций. Сказанное относится и к более широкому классу билинейных аппроксимаций, для построения или оценки несбалансированности которых, по-видимому, не предлагались ранее полиномиальные алгоритмы.

В настоящей работе описан вероятностный алгоритм, позволяющий оценивать сверху максимальную несбалансированность (в заданном широком классе, см. ниже формулу (3)) билинейных аппросимаций булевых отображений n переменных за время, линейно зависящее от n. Предложенный алгоритм базируется на развитии идеи, лежащей в основе усовершенствованного алгоритма Левина [5, 6], и может рассматриваться как обобщение одного из этапов последнего на случай билинейных аппроксимаций булевых отображений.

1. Постановка задачи и основные результаты

Обозначим V_n пространство двоичных векторов длины $n, F_{m \times n}$ — множество матриц размера $m \times n$ над полем $F = \mathrm{GF}(2)$. Для любых $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in V_n$ положим $xy = x_1y_1 \oplus ... \oplus x_ny_n, x \oplus y = (x_1 \oplus y_1, ..., x_n \oplus y_n)$. Строки произвольной матрицы $U \in F_{m \times n}$ обозначим $U_1, ..., U_m$.

Пусть $g:V_n\to V_n$ — булево отображение, заданное с помощью оракула (некоторого алгоритма, позволяющего вычислять значение g(x) по произвольному входному значению $x\in V_n$; см., например, [6]); *— бинарная операция на множестве V_n . Назовем билинейной аппроксимацией (между входами и выходами) отображения g произвольную функцию вида

$$x \mapsto xAg(x) \oplus \alpha x \oplus \beta g(x), \ x \in V_n,$$
 (1)

где $A \in F_{n \times n}, \ \alpha \in V_n, \ \beta \in V_n \setminus \{0\}$. Число

$$l_*^{(g)}(A,\alpha,\beta) = 2^{-n} \sum_{k \in V_n} \left(2^{-n} \sum_{x \in V_n} (-1)^{xAg(x*k) \oplus \alpha x \oplus \beta g(x*k)} \right)^2$$
 (2)

назовем (средней) несбалансированностью указанной аппроксимации (относительно операции *). Далее будем отождествлять функцию (1) с упорядоченным набором (A, α, β) .

Для любого подпространства L векторного пространства V_n обозначим

$$l_*^{(g)}(L,\beta) = \max\{l_*^{(g)}(A,\alpha,\beta) : A_1, ..., A_n \in L, \ \alpha \in V_n\}$$
(3)

максимальную несбалансированность билинейных аппроксимаций (A, α, β) отображения g по всем векторам $\alpha \in V_n$ и $n \times n$ -матрицам A, строки которых принадлежат подпространству L. Требуется построить вероятностный алгоритм, вычисляющий для заданных $\varepsilon, \delta \in (0,1)$ статистическую верхнюю оценку параметра (3), то есть такое случайное значение $\theta_{L,\beta} \in (0,1)$, для которого

$$\mathbf{P}\{l_*^{(g)}(L,\beta) \leqslant \theta_{L,\beta} + \varepsilon\} \geqslant 1 - \delta. \tag{4}$$

Введем ряд дополнительных обозначений.

Для любого натурального t обозначим B_t совокупность непустых подмножеств множества $\{1,2,...,t\}$. Пусть $X_1,...,X_t,Y_1,...,Y_t$ и $K^{(S)}$ ($S\in B_t$)— независимые в совокупности случайные векторы с равномерным распределением на множестве V_n . Положим $X_S = \bigoplus_{i \in S} X_i, Y_S = \bigoplus_{i \in S} Y_i,$

$$g_{1,S} = g(X_S * K^{(S)}), \quad g_{2,S} = g(Y_S * K^{(S)}), \quad S \in B_t.$$
 (5)

.

Для любых $A \in F_{n \times n}, \ \alpha \in V_n, \ \beta \in V_n \setminus \{0\}$ зададим случайную величину

$$\xi(A, \alpha, \beta) = (2^t - 1)^{-1} \sum_{S \in B_t} (-1)^{(X_S A \oplus \beta)g_{1,S} \oplus (Y_S A \oplus \beta)g_{2,S} \oplus \alpha(X_S \oplus Y_S)}.$$
 (6)

Заметим, что на основании равенств (2), (5), (6)

$$\mathbf{E}\,\xi(A,\alpha,\beta) = l_*^{(g)}(A,\alpha,\beta). \tag{7}$$

Кроме того, для любых различных множеств $S, S' \in B_t$ случайные векторы $(X_S, Y_S, K^{(S)})$ и $(X_{S'}, Y_{S'}, K^{(S')})$ независимы. Следовательно,

$$\mathbf{D}\xi(A,\alpha,\beta) = (2^{t} - 1)^{-2} \sum_{S \in B_{t}} \mathbf{D}\left((-1)^{(X_{S}A \oplus \beta)g_{1,S} \oplus (Y_{S}A \oplus \beta)g_{2,S} \oplus \alpha(X_{S} \oplus Y_{S})}\right) \leqslant$$

$$\leqslant (2^{t} - 1)^{-2}(2^{t} - 1)\mathbf{E}(1) = (2^{t} - 1)^{-1}.$$
(8)

Наконец, для любых $U, V \in F_{t \times n}, \beta \in V_n \setminus \{0\}$ положим

$$\eta_{(U, V, \beta)}^{a} = (2^{t} - 1)^{-1} \sum_{S \in B_{t}} (-1)^{(U_{S} \oplus \beta)g_{1,S} \oplus (V_{S} \oplus \beta)g_{2,S} \oplus a_{S}}, a = (a_{1}, ..., a_{t}) \in V_{t},$$
(9)

где $U_S = \bigoplus_{i \in S} U_i; \ V_S = \bigoplus_{i \in S} V_i; \ a_S = \bigoplus_{i \in S} a_i; \ S \in B_t$. Отметим, что вектор, составленный из значений (9), является (с точностью до сомножителя $(2^t-1)^{-1}$) произведением матрицы Адамара $H_t = ((-1)^{xy})_{x,y \in V_t}$ на вектор с координатами

$$g_{(U,V,\beta)}^S = \begin{cases} (-1)^{(U_S \oplus \beta)g_{1,S} \oplus (V_S \oplus \beta)g_{2,S}}, \text{ если } S \in B_t, \\ 0, \text{ если } S = \varnothing. \end{cases}$$
(10)

Утверждение 1. Пусть

$$\theta_{L,\beta} = \max\{\eta_{(U,V,\beta)}^a: U_1, ..., U_t, V_1, ..., V_t \in L, a \in V_t\}.$$
(11)

Тогда для любых $\varepsilon, \delta \in (0,1)$, удовлетворяющих условию

$$\delta^{-1}\varepsilon^{-2} \leqslant 2^t - 1,\tag{12}$$

справедливо неравенство (4).

Доказательство. Обозначим A^* и α^* соответственно матрицу A и вектор α , для которых достигается максимум в правой части равенства (3): $l_*^{(g)}(L,\beta) = l_*^{(g)}(A^*,\alpha^*,\beta)$. Положим $U_i^* = X_i A^*, \ V_i^* = Y_i A^*, \ a_i^* = \alpha^*(X_i \oplus Y_i), \ i = 1,\ldots,t; \ a^* = (a_1^*,\ldots,a_t^*);$ обозначим U^* и V^* случайные матрицы, составленные из вектор-строк U_1^*,\ldots,U_t^* и V_1^*,\ldots,V_t^* соответственно.

На основании формул (6), (9) справедливо равенство $\xi(A^*,\alpha^*,\beta) = \eta^{a^*}_{(U^*,V^*,\beta)}$. При этом в силу (11) событие $\{l_*^{(g)}(L,\beta) > \theta_{L,\beta} + \varepsilon\}$ влечет событие $\{\eta^{a^*}_{(U^*,V^*,\beta)} < l_*^{(g)}(A^*,\alpha^*,\beta) - \varepsilon\}$. Отсюда, используя соотношения (7), (8) и неравенство Чебышева, получим

$$\mathbf{P}\{l_*^{(g)}(L,\beta) > \theta_{L,\beta} + \varepsilon\} \leq \mathbf{P}\{\eta_{(U^*,V^*,\beta)}^{a^*} < l_*^{(g)}(A^*,\alpha^*,\beta) - \varepsilon\} =$$

$$= \mathbf{P}\{\xi(A^*,\alpha^*,\beta) < \mathbf{E}\xi(A^*,\alpha^*,\beta) - \varepsilon\} \leq \mathbf{P}\{|\xi(A^*,\alpha^*,\beta) - \mathbf{E}\xi(A^*,\alpha^*,\beta)| > \varepsilon\} \leq$$

$$\leq \varepsilon^{-2}\mathbf{D}\xi(A^*,\alpha^*,\beta) \leq \varepsilon^{-2}(2^t - 1)^{-1} \leq \delta,$$

где последнее неравенство вытекает из формулы (12).

Итак, при выполнении условия (12) справедливо неравенство (4), что и требовалось доказать. ■

Полученное утверждение позволяет предложить следующий вероятностный алгоритм вычисления верхних грании, параметра (3) по указанным выше исходным данным $g, \beta, L, \varepsilon, \delta$.

1. Положить

$$t = \lceil \log(1 + \delta^{-1} \varepsilon^{-2}) \rceil, \tag{13}$$

сгенерировать независимые в совокупности случайные векторы X_i , Y_i , $K^{(S)}$ ($i=1,\ldots,t,\ S\in B_t$) с равномерным распределением на множестве V_n , вычислить значения (5).

- 2. Для каждой пары матриц $U, V \in F_{t \times n}$, таких, что $U_i, V_i \in L \ (i = 1, \dots, t)$:
 - вычислить значения (10);
 - вычислить значения (9), применяя к вектору с координатами (10) алгоритм быстрого преобразования Адамара;
 - положить $\theta_{U,V,\beta} = \max\{\eta^a_{(U,V,\beta)} : a \in V_t\}.$
- 3. Положить $\theta_{L,\beta} = \max\{\theta_{U,V,\beta}: U_i, V_i \in L \ (i = 1, ..., t)\}.$

Обозначим $t^*(n)$ временную сложность операции *, т. е. максимальное число двоичных операций, выполняемых при вычислении значений x * y для любых $x, y \in V_n$.

Утверждение 2. Пусть dim L=r, где $r\equiv {\rm const}$ (не зависит от $n,\,\varepsilon,\,\delta$). Тогда временная сложность описанного алгоритма составляет

$$T_{\varepsilon,\delta}(n,r) = \mathcal{O}((\varepsilon^{-2}\delta^{-1})^{2r+1}(n\log(\varepsilon^{-2}\delta^{-1}) + \log^2(\varepsilon^{-2}\delta^{-1}) + t^*(n)))$$
(14)

двоичных операций и $O(\varepsilon^{-2}\delta^{-1})$ обращений к оракулу g. При этом объем памяти, необходимой для выполнения алгоритма, равен

$$M_{\varepsilon,\delta}(n) = \mathcal{O}(\varepsilon^{-2}\delta^{-1}(n + \log(\varepsilon^{-2}\delta^{-1})))$$
 бит. (15)

Доказательство. На шаге 1 для нахождения векторов X_S , Y_S ($S \in B_t$) достаточно выполнить $O(2^t nt)$ сложений по модулю 2. Следовательно, вычисление значений (5) можно осуществить за $T_1 = O(2^t nt + 2(2^t - 1)t^*(n)) = O(2^t (nt + t^*(n)))$ двоичных операций и $O(2^t)$ обращений к оракулу g.

На шаге 2 для каждой пары матриц U,V нахождение значений (10) потребует выполнения $O(2^tnt)$ двоичных операций, а вычисление значений (9) с помощью алгоритма быстрого преобразования Адамара (без учета деления на $2^t-1)-O(2^tt)$ операций сложения или вычитания целых чисел (см., например, [11], следствие 5.34). Поскольку разрядность указанных чисел не превосходит t, то двоичная временная сложность нахождения максимального значения (9) при фиксированных U,V и β не превосходит $O(2^tt(n+t))$. Наконец, так как число пар (U,V), удовлетворяющих условию $U_i,V_i\in L$ ($i=1,\ldots,t$), равно 2^{2rt} , то суммарная временная сложность второго и третьего шагов алгоритма составляет $T_2=O(2^{t(2r+1)}t(n+t))$ двоичных операций. Складывая выражения T_1 и T_2 , с учетом формулы (13) и условия $r\equiv$ const получим равенство (14).

Для оценки емкостной сложности алгоритма заметим, что объем памяти, необходимой для хранения чисел (5), составляет $O(2^t n)$ бит. Далее, для нахождения значения $\theta_{L,\beta}$ достаточно хранить текущие значения матриц U и V, соответствующие им

векторы (9), (10), а также ранее вычисленное значение $\theta_{\tilde{U},\tilde{V},\beta}$, соответствующее матрицам \tilde{U} и \tilde{V} , выбранным на предыдущем шаге вычислений. Суммарный объем необходимой для этого памяти не превосходит $O(nt+2^tt)$ бит, откуда следует справедливость формулы (15). \blacksquare

Отметим, что в практически значимом случае, когда $t^*(n) = O(n)$ и число t вида (13) меньше n, оценки (14), (15) упрощаются и принимают следующий вид:

$$T_{\varepsilon,\delta}(n,r) = \mathcal{O}(n(\varepsilon^{-2}\delta^{-1})^{2r+1}\log(\varepsilon^{-2}\delta^{-1})), M_{\varepsilon,\delta}(n) = \mathcal{O}(n\varepsilon^{-2}\delta^{-1}).$$

При $L=\{0\}$, $*=\oplus$ предложенный алгоритм позволяет оценивать свеху (с точностью ε и достоверностью $1-\delta$) максимум квадратов нормированных коэффициентов Уолша — Адамара булевой функции $f(x)=\beta g(x), x\in V_n$ за $O(n\varepsilon^{-2}\delta^{-1}\log(\varepsilon^{-2}\delta^{-1}))$ двочиных операций. В этом случае предложенный алгоритм по существу совпадает с первым этапом усовершенствованного алгоритма Левина [5, 6]. Отметим, что последний алгоритм формирует случайный список, содержащий с вероятностью не менее $1-\delta$ каждую аффинную функцию, находящуюся от функции f на расстоянии не более $2^{n-1}(1-\varepsilon)$, со сложностью $O(n\varepsilon^{-2}\delta^{-1}\log n\log(\varepsilon^{-2}\delta^{-1}))$ операций над n-разрядными целыми числами.

2. Результаты моделирования алгоритма

Описанный алгоритм был применен к исследованию отображений $g:V_{32}\to V_{32}$, построенных по схеме блока подстановки алгоритма шифрования ГОСТ 28147-89 [12]. Вычислительные эксперименты проводились для различных наборов узлов замены $s_i:V_4\to V_4,\ i=0,\ldots,7$, задающих отображение g, векторов $\beta\in V_{32}$ и подпространств L размерности 0 или 1. В качестве типового примера, иллюстрирующего полученные результаты, приведем оценки параметра (3), полученные для подстановки $g=(s_0,\ldots,s_7)$, операции * сложения по модулю 2^{32} на множестве V_{32} , вектора

и подпространства L, порожденного вектором

$$z = (0\ 0\ 1\ 0\ |\ 0\ 0\ 0\ 0\ |\ 0\ 0\ 0\ 0\ |\ 0\ 0\ 0\ 0\ |\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0).$$

Узлы замены $s_0, ..., s_7$ в выражении подстановки g определяются по табл. 1. Они характеризуются наименьшими значениями параметров

$$d(s) = \max \left\{ 2^{-4} \left| \left\{ x \in V_4 : s(x \oplus a) \oplus s(x) = b \right\} \right| : a, b \in V_4 \setminus \{0\} \right\},$$
$$l(s) = \max \left\{ \left(2^{-4} \sum_{x \in V_4} (-1)^{ax \oplus bs(x)} \right)^2 : a, b \in V_4 \setminus \{0\} \right\}$$

среди всех подстановок $s: V_4 \to V_4$ ($d(s_i) = l(s_i) = 0.25$ для любого i = 0, ..., 7) и рекомендуются в [13] для применения в алгоритме шифрования ГОСТ 28147-89.

В табл. 2 приведены численные оценки параметра (3). Отметим, что на основании следствий 1 и 3 работы [3] точное значение этого параметра в рассматриваемом случае может быть найдено по формуле

$$l(s_0, \beta_0) = \max_{(A, \alpha)} \left\{ 2^{-4} \sum_{k \in V_4} \left(2^{-4} \sum_{x \in V_4} \left(-1 \right)^{(xA s_0(x + k) \oplus \alpha x \oplus \beta_0 s_0(x + k)))} \right)^2 \right\},$$

	0	1	2	3	4	5	6	7	8	9	A	В	С	D	E	F
s_0	0	1	В	D	9	Е	6	7	С	5	8	3	F	2	4	A
s_1	0	1	2	4	3	5	8	A	7	9	6	D	В	E	С	F
s_2	0	1	В	2	8	6	F	3	E	A	4	9	D	5	7	С
s_3	0	1	В	2	8	3	F	6	E	A	4	9	D	5	7	С
s_4	0	4	В	2	8	6	A	1	E	F	3	9	D	5	7	С
s_5	0	4	В	2	8	3	F	1	E	A	6	9	D	5	7	С
s_6	0	В	F	9	1	5	6	8	3	A	4	С	Е	D	7	2
s_7	0	7	A	Е	9	1	D	8	С	2	В	F	3	5	4	6

Таблица 1 Набор «экстремальных» узлов замены [13]

где $\beta_0 = (1,0,0,0)$ — подвектор вектора (16), соответствующий подстановке s_0 ; x + k — сумма по модулю 2^4 двоичных целых чисел, соответствующих векторам $x, k \in V_4$; максимум берется по всем векторам $\alpha \in V_4$ и матрицам $A \in F_{4\times 4}$, строки которых принадлежат множеству $\{(0,0,0,0),(0,0,1,0)\}$. Таким образом, точное значение параметра $(3) - l(s_0,\beta_0) = 0.312500$, что отличается от его верхних оценок в среднем на 0.06 (см. последнюю колонку табл. 2).

Таблица 2 Результаты выполнения алгоритма $(\varepsilon=0,2,\ \delta=0,1,\ t=8)$

№ эксперимента	$\theta_{L,eta}$	$\theta_{L,\beta} + \varepsilon$
1	0,184314	0,384314
2	0,254902	0,454902
3	0,160784	0,360784
4	0,160784	0,360784
5	0,137255	0,337255
6	0,192157	0,392157
7	0,184314	0,384314
8	0,152941	0,352941
9	0,137255	0,337255
10	0,168627	0,368627

Для повышения точности оценок следует уменьшить значение ε (увеличить значение t) при применении алгоритма, что, очевидно, приведет к повышению его трудоемкости. При t=8 время работы компьютерной программы для ЭВМ Intel Pentium Dual-Core T4300 (2,1 ГГц, 3 Гбайт RAM) с использованием пакета прикладных программ Maple 13 составляет около 26 ч.

В целом, предложенный алгоритм представляется достаточно перспективным для криптографических приложений, прежде всего, для анализа и обоснования стойкости блочных шифров относительно билинейных атак.

ЛИТЕРАТУРА

- 1. Courtois N. T. Feistel schemes and bi-linear cryptanalysis // Advances in Cryptology CRYPTO'04. Springer Verlag, 2004. P. 23–40.
- 2. Алексейчук А. Н., Шевцов А. С. Показатели и оценки стойкости блочных шифров относительно статистических атак первого порядка // Реєстрація, зберігання і обробка даних. 2006. Т. 8. № 4. С. 53–63.

- 3. *Алексейчук А. Н., Шевцов А. С.* Верхние оценки несбалансированности билинейных аппроксимаций раундовых функций блочных шифров // Кибернетика и системный анализ. 2010. № 4. С. 42–51.
- 4. Goldreich O. and Levin L. A. A hard core predicate for all one-way functions // Proc. of the 21th ACM Sympos. of Theory of Computing. NY, USA: ACM, 1989. P. 25–32.
- 5. Levin L. A. Randomness and non-determinism // J. Symbolic Logic. 1993. V. 58. No. 3. P. 1102–1103.
- 6. Bshouty N., Jackson J., and Tamon C. More efficient PAC-learning of DNF with membership queries under the uniform distribution // Proc. 12th Annual Conf. on Comput. Learning Theory. NY, USA: ACM, 1999. P. 286–295.
- 7. Goldreich O., Rubinfeld R., and Sudan M. Learning polynomials with queries: the highly noisy case // SIAM J. Discrete Math. 2000. V.13. No. 4. P. 535-570. Extended version: http://people.csail.mit.edu/madhu/papers.html.
- 8. Kabatiansky G. and Tavernier C. List decoding of Reed-Muller codes // Proc. Ninth Int. Workshop on Algebraic and Comb. Coding Theory. Kranevo, Bulgaria, 2004. P. 230–235.
- 9. Trevisan L. Some applications of coding theory in computational complexity // http://eprint.arXiv:cs./0409044v1. 24 Sept., 2004.
- 10. Fourquet R., Loidreau P., and Tavernier C. Finding good linear approximations of block ciphers and its application to cryptanalysis of redused round DES // Proc. of the WCC 2009: ced.tavernier.free.fr/publications.
- 11. Логачев О. А., Сальников А. А., Ященко В. В. Булевы функции в теории кодирования и криптологии. М.: МЦНМО, 2004. 470 с.
- 12. ГОСТ 28147-89. Системы обработки информации. Защита криптографическая. Алгоритм криптографического преобразования. М.: Госстандарт СССР, 1989.
- 13. Ростовцев А. Γ ., Маховенко Е. Б. Введение в теорию итерированых шифров. СПб.: НПО «Мир и Семья», 2003. 302 с.