УДК 621.391:519.2

А.Н. АЛЕКСЕЙЧУК, С.Н. КОНЮШОК

УСОВЕРШЕНСТВОВАННЫЙ ТЕСТ *k*-МЕРНОСТИ ДЛЯ БУЛЕВЫХ ФУНКЦИЙ

Ключевые слова: *проверка свойств булевых функций, вероятностный алгоритм, k-мерная функция, преобразование Уолша–Адамара.*

Настоящая статья посвящена решению задачи тестирования свойства k-мерности булевых функций. Предложен вероятностный алгоритм (тест k-мерности), имеющий лучшие характеристики эффективности по сравнению с ранее известным тестом [1].

Для уточнения постановки задачи приведем определения основных понятий и сформулируем вспомогательные утверждения о свойствах булевых функций. (Более подробную информацию по этим вопросам можно найти в [2].)

Обозначим V_n векторное пространство размерности n над полем $F = \mathbf{GF}(2)$. Сумма векторов $\alpha = (\alpha_1, \dots, \alpha_n)$, $x = (x_1, \dots, x_n) \in V_n$ определяется по формуле $\alpha \oplus x = (\alpha_1 \oplus x_1, \dots, \alpha_n \oplus x_n)$, а булево скалярное произведение — по формуле $\alpha x = \alpha_1 x_1 \oplus \dots \oplus \alpha_n x_n$ (здесь и далее символ $\oplus$ обозначает операцию сложения как элементов поля F, так и векторов над этим полем). Для любого множества $M \subseteq V_n$ обозначим $M^{\perp}$ подпространство векторного пространства V_n, дуальное к M: $M^{\perp} = \{\alpha \in V_n |\ \forall\, x \in M : \alpha x = 0\}$; для любых $a, b \in \mathbf{Z}$ положим $\overline{a, b} = \{i \in \mathbf{Z}: a \le i \le b\}$.

Обозначим B_n множество булевых функций от n переменных. Расстояние Хэмминга между функциями $f, g \in B_n$ определяется по формуле $d(f, g) = |\{x \in V_n : f(x) \ne g(x)\}|$, а расстояние от функции $f \in B_n$ до множества $U \subseteq B_n$ — по формуле $d(f, U) = \min\limits_{g \in U} d(f, g)$.

Для любой функции $f \in B_n$ положим

$$\hat{f}(\alpha) = 2^{-n} \sum_{x \in V_n} (-1)^{f(x) \oplus \alpha x},\ \alpha \in V_n; \tag{1}$$

$$I_f = \{\alpha \in V_n |\ \forall\, x \in V_n : f(x \oplus \alpha) = f(x)\}, \tag{2}$$

$$Sp(f) = \{\alpha \in V_n | \hat{f}(\alpha) \ne 0\}. \tag{3}$$

Числа (1) называются нормированными коэффициентами Уолша–Адамара функции f; при этом обычные (ненормированные) коэффициенты Уолша–Адамара определяются по формуле $W_f(\alpha) = 2^n \hat{f}(\alpha)$, $\alpha \in V_n$ [2]. Известно (см. теорему 1 в [3] или задачу 2.112 в [2]), что для любой функции $f \in B_n$ выполняется равенство

$$I_f = Sp(f)^{\perp}. \tag{4}$$

Функция $f \in B_n$ называется k-мерной [1], $k \in \overline{0, n-1}$, если множество (3) порождает в V_n подпространство размерности не более k или, что равносильно, если существует не менее $n - k$ линейно независимых векторов, принадлежащих множеству (2).

Обозначим $B_n(k)$ множество всех k-мерных функций от n переменных, $k \in \overline{0, n-1}$. Известно, что при умеренных по сравнению с n значениях k функции, близкие к k-мерным, обладают криптографическими слабостями, что позволяет осуществлять некоторые атаки на генераторы гаммы, построенные на основе указанных функций [4–6]. В связи с этим важной задачей является разработка эффективных алгоритмов проверки свойства k-мерности булевых функций.

Отметим, что если функция $f \in B_n$ задана с помощью вектора значений (таблицы истинности), то для проверки условия $f \in B_n(k)$ можно применить естественный детерминированный алгоритм, трудоемкость которого составляет $O(n^2 2^n)$ двоичных операций. Этот алгоритм состоит в вычислении всех значений (1) с помощью быстрого преобразования Адамара (см., например, [2], с. 217), построении множества (3) и нахождении базиса векторного пространства (4) методом Гаусса. Функция f является k-мерной тогда и только тогда, когда полученный базис содержит не менее $n-k$ векторов. Ясно, что этот алгоритм не применим на практике, если n является достаточно большим числом (например, $n \geq 64$), а функция f задается с помощью оракула (некоторого алгоритма, позволяющего вычислять значения $f(x)$ по произвольным входным аргументам $x \in V_n$).

В [1] предложен вероятностный алгоритм или тест k-мерности, который для любой функции $f \in B_n$, заданной с помощью оракула, и чисел $k \in \overline{0, n-1}$, $\varepsilon \in (0,1)$ проверяет основную гипотезу H_0: $f \in B_n(k)$ против альтернативы H_1: $d(f, B_n(k)) \geq 2^n \varepsilon$. Этот алгоритм состоит в генерации независимых случайных равновероятных векторов $h_1, \ldots, h_l \in V_n$ и проверке равенств

$$f(h_j \oplus Z_{ij}) = f(Z_{ij}),\ i \in \overline{1, m}, \tag{5}$$

для каждого $j \in \overline{1, l}$, где Z_{ij} — независимые в совокупности и не зависящие от $h_1, \ldots, h_l$ случайные равновероятные векторы из V_n. Обозначим ν_l число значений $j \in \overline{1, l}$, для которых выполняются равенства (5). Тогда гипотеза H_0 принимается, если $\frac{\nu_l}{l} \geq 0{,}9 \cdot 2^{-k}$, и отклоняется в противном случае. В [1] предложено выбрать $l = 2^k C$, $m = 2^k k\, \varepsilon^{-1} C'$, где $C, C' = \text{const}$, что приводит к оценке трудоемкости алгоритма, составляющей $O(n 2^{2k} k\, \varepsilon^{-1})$ двоичных операций.

Для оценки вероятности ошибки первого рода (т.е. вероятности того, что тест «не признает» таковой k-мерную функцию) в [1] используется неравенство Чернова

$$\mathbf{P}\left(\frac{\nu_l}{l} < 0{,}9 \cdot 2^{-k} \mid H_0 \right) \leq \mathbf{P}\left(\frac{\nu_l}{l} - \mathbf{E}\frac{\nu_l}{l} < -0{,}1 \cdot 2^{-k} \mid H_0 \right) \leq \exp\left\{ -0{,}02 \cdot \frac{C}{2^k} \right\}. \tag{6}$$

Отметим, что выражение в правой части (6) зависит от k и не стремится к нулю, если k является (сколь угодно медленно) растущей функцией от n, например, $k = \lceil \log n \rceil$, $n \to \infty$.

Далее предложен более эффективный тест k-мерности, трудоемкость которого составляет $O(n 2^k k^2\, \varepsilon^{-1})$ двоичных операций. При этом верхняя граница вероятности ошибки первого рода предложенного теста не зависит от k, а верхняя граница вероятности ошибки второго рода по существу та же, что и для теста из [1]. Показано также, что при некотором естественном изменении альтернативы H_1 можно построить односторонний (с нулевой вероятностью ошибки первого рода) тест

k-мерности, трудоемкость которого составляет $O(n(2^k + k\varepsilon^{-2})\log(2^k + k\varepsilon^{-2}))$ двоичных операций.

Сформулируем основные результаты. Ключевая идея, лежащая в основе предлагаемого теста, состоит в том, чтобы не выбирать векторы $h_1, \dots, h_l$ наугад, а сформировать их с помощью вспомогательной процедуры таким образом, чтобы множество указанных векторов с высокой вероятностью содержалось во множестве I_f, если f — k-мерная функция. Для этого рассмотрим сужение функции f на случайно выбранное подпространство векторного пространства V_n. Отметим, что идея использовать такие сужения при проверке различных свойств булевых функций восходит, по-видимому, к работе Л. Левина [7] и лежит в основе ряда вероятностных алгоритмов тестирования степени полиномов от нескольких переменных над полем из двух элементов [8, 9]. В настоящей статье эта идея реализована следующим образом.

Обозначим $F_{m\times n}$ множество матриц размера $m\times n$ над полем F. Для любой матрицы $X \in F_{t\times n}$, где $k < t < n$, положим $f_X(u) = f(uX)$, $u \in V_t$.

Теорема 1. Если $f \in B_n$ — k-мерная функция, то функция f_X также является k-мерной. При этом вероятность того, что при случайном равновероятном выборе $t\times n$-матрицы X множество $\{aX : a \in I_{f_X}\}$ содержится во множестве I_f, больше либо равна $1-2^{k-t}$.

Алгоритм проверки k-мерности булевых функций, основанный на теореме 1, имеет следующий вид.

Алгоритм 1

Исходные данные: $f \in B_n$, $k \in \overline{0, n-1}$, $\varepsilon \in (0,1)$.

Параметры: $t = k + c$, $m = 2^{t+4} t\, \varepsilon^{-1}\delta^{-1}$, где $c \in \mathbf{N}$, $\delta \in (0, 1/2)$, $c, \delta = \text{const}$.

Шаг 1. Сгенерировать случайную равновероятную $t\times n$-матрицу X, построить множество $Sp(f_X)$, по которому найти базис $a_1, \dots, a_l$ векторного пространства $I_{f_X} = Sp(f_X)^{\perp}$. Проверить условие

$$l \geq t - k, \tag{7}$$

при выполнении которого перейти к шагу 2. В противном случае принять гипотезу H_1: $d(f, B_n(k)) \geq 2^n \varepsilon$.

Шаг 2. Для каждого $j \in \overline{1, l}$ положить $h_j = a_j X$, сгенерировать независимые случайные равновероятные векторы $Z_{1j}, \dots, Z_{mj}$ и проверить равенства (5). При выполнении указанных равенств для всех $j \in \overline{1, l}$ принять гипотезу H_0: $f \in B_n(k)$; в противном случае принять гипотезу H_1.

Теорема 2. Описанный алгоритм выполняет $O(2^k k^2 \varepsilon^{-1})$ запросов к оракулу f, а его трудоемкость составляет $O(n 2^k k^2 \varepsilon^{-1})$ двоичных операций. При этом вероятность ошибки первого рода (отвергнуть истинную гипотезу H_0) не превышает 2^{-c}, а вероятность ошибки второго рода (отвергнуть истинную гипотезу H_1) не превышает $\max\{5\cdot 2^{-c-1}, \delta + \exp\{-7c2^c\}\}$.

Пусть теперь требуется проверить гипотезу H_0: $f \in B_n(k)$ против альтернативы K, состоящей в том, что множество $\{\alpha \in V_n : |\hat{f}(\alpha)| \geq \varepsilon\}$ порождает подпространство размерности не менее $k+1$. В этом случае можно предложить односторонний тест, по существу состоящий в выполнении шага 1 алгоритма 1.

Алгоритм 2

Исходные данные: $f \in B_n$, $k \in \overline{0, n-1}$, $\varepsilon \in (0,1)$.

Параметр: $t = \lceil \log(1 + \delta^{-1}(2^{k+1} + 4(k+1)\varepsilon^{-2})) \rceil$, где $\delta \in (0, 1/2)$, $\delta = \text{const}$.

Сгенерировать случайную равновероятную $t \times n$-матрицу X. Если f_X — k-мерная функция, принять гипотезу H_0; в противном случае принять гипотезу K.

Теорема 3. Описанный алгоритм выполняет $O(\delta^{-1}(2^k + k\varepsilon^{-2}))$ запросов к оракулу f, а его трудоемкость составляет $O(n\delta^{-1}(2^k + k\varepsilon^{-2})\log(\delta^{-1}(2^k + k\varepsilon^{-2})))$ двоичных операций. При этом вероятность ошибки первого рода равна нулю, а вероятность ошибки второго рода не превышает δ.

Доказательство теоремы 1. Установим ряд вспомогательных свойств k-мерных булевых функций. Следующая далее лемма по существу совпадает с утверждением 2 в работе [5].

Лемма 1. Функция $f \in B_n$ является k-мерной тогда и только тогда, когда существуют число $r \in \overline{0,k}$, матрица $A \in F_{n\times r}$ и функция $g \in B_r$ такие, что

$$f(x) = g(xA),\ x \in V_n. \tag{8}$$

Если при этом r — наименьшее число с указанным свойством, то $I_f = \{\alpha \in V_n: \alpha A = 0\}$ и $\dim I_f = n - r$.

Назовем представление k-мерной функции f в виде (8), соответствующее наименьшему возможному значению $r \in \overline{0,k}$, неприводимым представлением этой функции.

Следствие 1. Представление (8) является неприводимым тогда и только тогда, когда $\operatorname{rank} A = r$ и $I_g = \{0\}$.

Лемма 2. Пусть (8) — неприводимое представление k-мерной функции $f \in B_n$, где $g \in B_r$, $r \in \overline{0,k}$. Тогда для любой матрицы $X \in F_{t\times n}$, где $k < t < n$, функция f_X является k-мерной. Более того, если $\operatorname{rank} XA = r$, то

$$\{aX: a \in I_{f_X}\} \subseteq I_f. \tag{9}$$

Доказательство. Из равенства (8) вытекает, что $f_X(u) = f(uX) = g(u(XA))$, $u \in V_t$. Следовательно, f_X является k-мерной функцией согласно лемме 1.

Пусть $\operatorname{rank} XA = r$. Поскольку (8) — неприводимое представление функции f, на основании следствия 1 справедливо равенство $I_g = \{0\}$. Следовательно, $f_X(u) = g(u(XA))$, $u \in V_t$ есть неприводимое представление функции f_X, откуда согласно лемме 1 вытекает, что $I_{f_X} = \{a \in V_t: aXA = 0\}$. Таким образом, если $a \in I_{f_X}$, то для любого $z \in V_n$ выполняются равенства $f(aX \oplus z) = g(aXA \oplus zA) = g(zA) = f(z)$, т.е. $aX \in I_f$, что и требовалось доказать.

Лемма 3. Пусть $\alpha_1, \ldots, \alpha_r \in V_n$ — линейно независимые векторы, $r \le t < n$. Тогда вероятность того, что при случайном равновероятном выборе матрицы $X \in F_{t\times n}$ векторы $X\alpha_1, \ldots, X\alpha_r$ являются линейно зависимыми, не превышает 2^{r-t}.

Доказательство. Если векторы $X\alpha_1, \ldots, X\alpha_r$ линейно зависимы, то существует ненулевой вектор $\alpha = c_1\alpha_1 \oplus \cdots \oplus c_r\alpha_r$ ($c_i \in F$, $i \in \overline{1,r}$) такой, что $X\alpha = 0$. Вероятность последнего события равна 2^{-t}. Следовательно, вероятность того, что векторы $X\alpha_1, \ldots, X\alpha_r$ линейно зависимы, не превышает $(2^r - 1)2^{-t}$. Лемма доказана.

Исходя из лемм 2, 3, нетрудно убедиться в справедливости теоремы 1. Действительно, рассмотрим неприводимое представление k-мерной функции f в виде (8), где $g \in B_r$, $r \in \overline{0,k}$. Согласно лемме 2 при случайном равновероятном выборе матрицы $X \in F_{t\times n}$ вероятность события (9) не меньше вероятности события $\{\operatorname{rank} XA = r\}$. Последняя, в свою очередь, больше либо равна $1 - 2^{r-t} \ge 1 - 2^{k-t}$ на основании леммы 3.

Теорема 1 доказана.

Доказательство теоремы 2. Сформулируем два вспомогательные утверждения, первое из которых представляет собой «факт 11» из [1], а второе — вариант неравенства Чебышева (см. утверждение 4 в [1]).

Лемма 4. Пусть Z — случайный равновероятный вектор со значениями во множестве V_n. Тогда для любых $f \in B_n$, $y \in V_n$ выполняется равенство

$$\mathbf{P}_Z \{f(y \oplus Z) \neq f(Z)\} = \sum_{\substack{\alpha \in V_n: \\ y\alpha = 1}} |\hat{f}(\alpha)|^2.$$

Лемма 5. Пусть $\xi = \sum_{i=1}^{N} \xi_i$, где $\xi_1, \ldots, \xi_N$ — попарно независимые случайные величины такие, что $0 \le \xi_i \le \tau$, $i \in \overline{1, N}$. Тогда если $\mathbf{E}\xi > 0$, то для любого $\delta > 0$ справедливо неравенство $\mathbf{P}\{\xi \le (1-\delta)\mathbf{E}\xi\} \le \dfrac{\tau}{\delta^2 \mathbf{E}\xi}$.

Лемма 6. Алгоритм 1 характеризуется вероятностью ошибки первого рода не более 2^{-c} и выполняет $O(2^k k^2 \varepsilon^{-1})$ запросов к оракулу f. При этом его трудоемкость составляет $O(n 2^k k^2 \varepsilon^{-1})$ двоичных операций.

Доказательство. Первое утверждение леммы следует из теоремы 1. Действительно, если f является k-мерной функцией, то такой же является функция f_X. Следовательно, неравенство (7) заведомо выполняется, и алгоритм может совершить ошибку только в том случае, когда на шаге 2 нарушается хотя бы одно из равенств (5). Однако на основании теоремы 1 вероятность последнего события не превышает $2^{k-t} = 2^{-c}$, что и требовалось доказать.

Оценим трудоемкость алгоритма. На шаге 1 для вычисления значений функции f_X необходимо выполнить 2^t запросов к оракулу f, каждый из которых требует $O(nt)$ двоичных операций. Далее, для нахождения коэффициентов Уолша–Адамара функции f_X достаточно выполнить $O(2^t t)$ сложений или вычитаний не более чем t-разрядных целых чисел (см., например, следствие 5.34 из [2]), что составляет $O(2^t t^2)$ двоичных операций. Такое же время потребуется для построения базиса векторного пространства I_{f_X} с помощью метода Гаусса. На шаге 2 проверка равенств (5) для полученных базисных векторов потребует $2ml \le 2mt$ запросов к оракулу f, что составляет $O(nmt)$ двоичных операций.

Таким образом, с учетом указанных выше значений параметров m и t общее число запросов к оракулу равно $O(2^t + mt) = O(2^k k^2 \varepsilon^{-1})$, а трудоемкость алгоритма составляет $O(n2^t t + 2^t t^2 + nmt) = O(n2^k k^2 \varepsilon^{-1})$ двоичных операций.

Лемма доказана.

Для оценки вероятности ошибки второго рода воспользуемся методом, предложенным в [1]. Зафиксируем число $\theta \in (0,1)$ и рассмотрим множества

$$R(\theta) = \{\alpha \in V_n \setminus \{0\}: |\hat{f}(\alpha)| \ge \theta\}, \; S(\theta) = \{\alpha \in V_n \setminus \{0\}: |\hat{f}(\alpha)| < \theta\}.$$

Покажем сначала, что если множество $R(\theta)$ порождает пространство размерности более k (и, следовательно, f заведомо не является k-мерной функцией), то алгоритм 1 совершит ошибку с пренебрежимо малой вероятностью.

Лемма 7. Пусть функция f такова, что множество $R(\theta)$ содержит по крайней мере $k+1$ линейно независимых векторов $\alpha_1, \ldots, \alpha_{k+1}$. Тогда вероятность того, что алгоритм 1 совершит ошибку (т.е. посчитает f k-мерной функцией), не превышает

$$p_1 = 2^{1-c} + (k+1)(1-\theta^2)^m 2^{k+c-1}. \qquad (10)$$

Доказательство. Пусть алгоритм совершает ошибку. Тогда либо векторы $X\alpha_1, \dots, X\alpha_{k+1}$ линейно зависимы (согласно лемме 3 вероятность этого события не превышает $2^{k+1-t} = 2^{1-c}$), либо они линейно независимы, и тогда как минимум один из них, скажем, $X\alpha_i$, $i \in \overline{1, k+1}$, не принадлежит множеству $Sp(f_X)$. Поскольку $I_{f_X} = Sp(f_X)^{\perp}$, по крайней мере один из базисных векторов $a_1, \dots, a_l$ пространства I_{f_X} не ортогонален вектору $X\alpha_i$. Следовательно, существует ненулевой вектор $a \in V_t$ такой, что $aX\alpha_i = 1$, и для вектора $h_j = aX$ выполняются равенства (5). Таким образом, вероятность ошибки алгоритма не превышает

$$2^{1-c} + \mathbf{P}_{X, Z_1, \dots, Z_m}\left(\bigcup_{i=1}^{k+1} \bigcup_{a \in V_t \setminus \{0\}} \{aX\alpha_i = 1,\ f(aX \oplus Z_s) = f(Z_s),\ s \in \overline{1, m}\} \right) \le$$

$$\le 2^{1-c} + (k+1)2^t \max_{\substack{i \in \overline{1, k+1}, \\ a \in V_t \setminus \{0\}}} \mathbf{P}_{X, Z_1, \dots, Z_m} \{aX\alpha_i = 1,\ f(aX \oplus Z_s) = f(Z_s),\ s \in \overline{1, m}\}.$$

Далее, в силу независимости и равновероятности случайной матрицы X и векторов $Z_1, \dots, Z_m$ для любых $i \in \overline{1, k+1}$, $a \in V_t \setminus \{0\}$ выполняется равенство

$$\mathbf{P}_{X, Z_1, \dots, Z_m} \{aX\alpha_i = 1,\ f(aX \oplus Z_s) = f(Z_s),\ s \in \overline{1, m}\} =$$

$$= \sum_{\substack{y \in V_n: \\ y\alpha_i = 1}} 2^{-nt} \sum_{\substack{X \in F_{t \times n}: \\ aX = y}} (\mathbf{P}_Z \{f(y \oplus Z) = f(Z)\})^m.$$

При этом, если $y\alpha_i = 1$, то на основании леммы 4 и условия $\alpha_i \in R(\theta)$ справедливы следующие соотношения:

$$\mathbf{P}_Z \{f(y \oplus Z) \ne f(Z)\} = \sum_{\substack{\alpha \in V_n: \\ y\alpha = 1}} |\hat{f}(\alpha)|^2 \ge |\hat{f}(\alpha_i)| \ge \theta^2.$$

Следовательно,

$$\mathbf{P}_{X, Z_1, \dots, Z_m} \{aX\alpha_i = 1,\ f(aX \oplus Z_s) = f(Z_s),\ s \in \overline{1, m}\} \le$$

$$\le \sum_{\substack{y \in V_n: \\ y\alpha_i = 1}} 2^{-nt} \sum_{\substack{X \in F_{t \times n}: \\ aX = y}} (1 - \theta^2)^m = \frac{1}{2}(1 - \theta^2)^m.$$

Из полученных неравенств вытекает, что вероятность ошибки алгоритма не превышает значения (10). Лемма доказана.

Отметим, что в приведенном доказательстве не используется предположение о том, что функция f находится на расстоянии не менее $2^n \varepsilon$ от множества $B_n(k)$. Остается рассмотреть случай, когда множество $R(\theta)$ порождает пространство размерности не более k, при этом рассуждения во многом близки к используемым в [1]. В частности, следующая далее лемма по существу совпадает с леммой 8 в [1].

Лемма 8. Пусть функция f находится на расстоянии не менее $2^n \varepsilon$ от множества $B_n(k)$, а множество $R(\theta)$ порождает подпространство размерности не более k. Тогда

$$\sum_{\alpha \in S(\theta)} |\hat{f}(\alpha)|^2 \ge \varepsilon. \tag{11}$$

Итак, для завершения доказательства теоремы 2 остается оценить вероятность ошибки алгоритма в предположении справедливости неравенства (11).

Лемма 9. Пусть выполняется неравенство (11). Тогда вероятность того, что алгоритм 1 совершит ошибку (т.е. посчитает f k-мерной функцией), не превышает

$$p_2 = 2^{k+c}(8\varepsilon^{-1}\theta^2 + (1-\varepsilon/4)^m). \tag{12}$$

Доказательство. Если алгоритм совершает ошибку, то существует вектор $a \in V_t \setminus \{0\}$, для которого случайный вектор $h_j = aX$ удовлетворяет равенствам (5). Поскольку вектор h_j имеет равномерное распределение на множестве V_n, вероятность ошибки алгоритма не превышает

$$2^t \mathbf{P}_{Y, Z_1, \dots, Z_m}\{f(Y \oplus Z_s) = f(Z_s),\ s \in \overline{1, m}\}, \tag{13}$$

где $Y, Z_1, \dots, Z_m$ — независимые в совокупности случайные равновероятные векторы со значениями в V_n.

Для оценки параметра (13) воспользуемся рассуждениями, аналогичными приведенным в доказательстве леммы 7 в [1]. Рассмотрим случайную величину

$$\xi(Y) = \sum_{\alpha \in S(\theta)} |f(\alpha)|^2 I_\alpha(Y),$$

где $I_\alpha(Y)$ — индикатор события $Y\alpha = 1$, $\alpha \in S(\theta)$. Поскольку согласно определению все векторы $\alpha \in S(\theta)$ являются ненулевыми, в силу неравенства (11) имеем

$$\mathbf{E}\xi(Y) = \frac{1}{2} \sum_{\alpha \in S(\theta)} |f(\alpha)|^2 \geq \frac{\varepsilon}{2}. \tag{14}$$

Кроме того, случайные величины $I_\alpha(Y)$, $\alpha \in S(\theta)$, являются попарно независимыми. Следовательно, на основании леммы 5, определения множества $S(\theta)$ и неравенства (14) получим, что

$$\mathbf{P}_Y\{\xi(Y) \leq \frac{1}{2}\mathbf{E}\xi(Y)\} \leq \frac{\max\limits_{\alpha \in S(\theta)} |\hat{f}(\alpha)|}{1/4 \cdot \mathbf{E}\xi(Y)} \leq 8\varepsilon^{-1}\theta^2.$$

Заметим теперь, что

$$\mathbf{P}_{Y, Z_1, \dots, Z_m}\{f(Y \oplus Z_s) = f(Z_s),\ s \in \overline{1, m}\} \leq \mathbf{P}_Y\left\{\xi(Y) \leq \frac{1}{2}\mathbf{E}\xi(Y)\right\} +$$

$$+ \mathbf{P}_{Y, Z_1, \dots, Z_m}\left\{f(Y \oplus Z_s) = f(Z_s),\ s \in \overline{1, m},\ \xi(Y) > \frac{1}{2}\mathbf{E}\xi(Y)\right\} \leq$$

$$\leq 8\varepsilon^{-1}\theta^2 + 2^{-n} \sum_{\substack{y \in V_n: \\ \xi(y) > \frac{1}{2}\mathbf{E}\xi(Y)}} (\mathbf{P}_Z\{f(y \oplus Z) = f(Z)\})^m.$$

При этом на основании леммы 4 и неравенства (14) для любого $y \in V_n$ такого, что $\xi(y) > \frac{1}{2}\mathbf{E}\xi(Y)$, справедливы следующие неравенства:

$$\mathbf{P}_Z\{f(y \oplus Z) \neq f(Z)\} = \sum_{\substack{\alpha \in V_n: \\ y\alpha = 1}} |\hat{f}(\alpha)|^2 \geq \xi(y) > \frac{\varepsilon}{4}.$$

Из последних двух соотношений получаем окончательную оценку параметра (13):

$$2^t \mathbf{P}_{Y, Z_1, \dots, Z_m} \{f(Y \oplus Z_s) = f(Z_s),\ s \in \overline{1, m}\} \le 2^t (8\varepsilon^{-1}\theta^2 + (1-\varepsilon/4)^m).$$

Таким образом, вероятность ошибки алгоритма не превышает значения (12), что и требовалось доказать.

Завершая доказательство теоремы 2, отметим, что на основании лемм 7 и 9 вероятность ошибки второго рода алгоритма 1 не превышает $\max\{p_1, p_2\}$, где

$$p_1 = 2^{1-c} + (k+1)(1-\theta^2)^m 2^{k+c-1},\ p_2 = 2^{k+c}(8\varepsilon^{-1}\theta^2 + (1-\varepsilon/4)^m). \quad (15)$$

Полагая в формулах (15) $\theta^2 = 2^{-t-3}\varepsilon\delta$, $m = 2^{t+4} t\, \varepsilon^{-1}\delta^{-1}$, где $\delta \in (0, 1/2)$, получаем, что

$$p_1 \le 2^{1-c} + 1/2 \cdot (k+1) \exp\{k + c - \theta^2 m\} =$$

$$= 2^{1-c} + 1/2 \cdot e^{-c}(k+1)e^{-k} < 2^{1-c} + 1/2 \cdot 2^{-c} = 5 \cdot 2^{-c-1},$$

$$p_2 = 2^{k+c}(8\varepsilon^{-1}\theta^2 + (1-\varepsilon/4)^m) = \delta + 2^{k+c}(1-\varepsilon/4)^m < \delta + 2^{k+c}\exp\{-m\varepsilon/4\} =$$

$$= \delta + 2^{k+c}\exp\{-4t2^t\delta^{-1}\} < \delta + 2^{k+c}\exp\{-8t2^t\} < \delta + \exp\{t - 8t2^t\} <$$

$$< \delta + \exp\{-7t2^t\} < \delta + \exp\{-7c2^c\}.$$

Итак, вероятность ошибки второго рода алгоритма 1 не превышает $\max\{5 \cdot 2^{-c-1},\ \delta + \exp\{-7c2^c\}\}$.

Теорема 2 полностью доказана.

Доказательство теоремы 3. Согласно теореме 1 вероятность ошибки первого рода алгоритма 2 равна нулю. Кроме того, повторяя рассуждения, приведенные в доказательстве леммы 6, получим с учетом значения параметра t, что число запросов к оракулу при выполнении алгоритма составляет $O(2^t) = O(\delta^{-1}(2^k + k\varepsilon^{-2}))$, а трудоемкость алгоритма равна

$$O(n2^t t + 2^t t^2) = O(n2^t t) = O(n\delta^{-1}(2^k + k\varepsilon^{-2})\log(\delta^{-1}(2^k + k\varepsilon^{-2})))$$

двоичных операций.

Покажем, что при выполнении условия теоремы 3 вероятность ошибки второго рода алгоритма не превышает δ.

Рассмотрим случайные величины

$$\eta_a(X) = \frac{1}{2^t - 1} \sum_{u \in V_t \setminus \{0\}} (-1)^{f(uX) \oplus ua},\ a \in V_t. \quad (16)$$

Известно [7] (и нетрудно проверить, опираясь на неравенство Чебышева и условие попарной независимости случайных векторов uX, $u \in V_t \setminus \{0\}$), что для любых $\alpha \in V_n$ и $\varepsilon \in (0, 1)$ справедливы следующие соотношения:

$$\mathbf{E}\eta_{X\alpha}(X) = \hat{f}(\alpha),$$

$$\mathbf{P}_X \{|\,|\eta_{X\alpha}(X)| - |\hat{f}(\alpha)|\,| \ge \varepsilon\} \le \frac{\varepsilon^{-2}}{2^t - 1}. \quad (17)$$

Пусть справедлива гипотеза K, и алгоритм совершает ошибку. Тогда множество $\{\alpha \in V_n : |\hat{f}(\alpha)| \ge \varepsilon\}$ содержит $k+1$ линейно независимых векторов

$\alpha_1, \ldots, \alpha_{k+1}$, а множество $Sp(f_X)$ порождает подпространство размерности не более k, поскольку f_X является k-мерной функцией. Следовательно, вероятность ошибки алгоритма не превышает суммы вероятностей двух событий: Ω_1 и Ω_2. Первое состоит в том, что векторы $X\alpha_1, \ldots, X\alpha_{k+1}$ линейно зависимы, а второе — что эти векторы линейно независимы, а множество $Sp(f_X)$ порождает подпространство размерности не более k.

На основании леммы 3 вероятность события Ω_1 не превышает 2^{k+1-t}. Кроме того, из определения множества $Sp(f_X)$ и формулы (16) вытекает равенство $Sp(f_X) = \{a \in V_t : \eta_a(X) \neq (2^t - 1)^{-1}(-1)^{\overline{f(0) \oplus 1}}\}$. Следовательно, если происходит событие Ω_2, то найдется число $i \in \overline{1, k+1}$ такое, что $X\alpha_i \notin Sp(f_X)$, и значит, $|\eta_{X\alpha_i}(X)| = (2^t - 1)^{-1}$. Наконец, поскольку в силу выбора значений t и δ

$$2^t - 1 \geq \delta^{-1}(2^{k+1} + 4(k+1)\varepsilon^{-2}) > 2\varepsilon^{-1},$$

то $|\hat{f}(\alpha_i)| \geq \varepsilon > 2(2^t - 1)^{-1}$. Отсюда с учетом неравенства (17) следует, что

$$\mathbf{P}_X\{|\eta_{X\alpha_i}(X)| = (2^t - 1)^{-1}\} \leq \mathbf{P}_X\{|\eta_{X\alpha_i}(X)| - |\hat{f}(\alpha_i)| \leq -(\varepsilon - (2^t - 1)^{-1})\} \leq$$

$$\leq \mathbf{P}_X\{||\eta_{X\alpha_i}(X)| - |\hat{f}(\alpha_i)|| \geq \varepsilon - (2^t - 1)^{-1}\} \leq \frac{(\varepsilon - (2^t - 1)^{-1})^{-2}}{2^t - 1} < \frac{4\varepsilon^{-2}}{2^t - 1}.$$

Таким образом, вероятность события Ω_2 не превышает $\dfrac{4\varepsilon^{-2}(k+1)}{2^t - 1}$, а вероятность ошибки второго рода не превышает

$$2^{k+1-t} + \frac{4\varepsilon^{-2}(k+1)}{2^t - 1} \leq \frac{2^{k+1} + 4\varepsilon^{-2}(k+1)}{2^t - 1} \leq \delta.$$

Теорема 3 доказана.

СПИСОК ЛИТЕРАТУРЫ

1. Testing Fourier dimensionality and sparsity / P. Gopalan, R. O'Donnell, A. Servedio et al. // SIAM J. Comput. — 2011. — **40**(4). — P. 1075–1100.
2. Логачев О.А., Сальников А.А., Ященко В.В. Булевы функции в теории кодирования и криптологии. — М.: МЦНМО, 2004. — 470 с.
3. Ященко В.В. О критерии распространения для булевых функций и бент-функций // Проблемы передачи информации. — 1997. — **33**, № 1. — С. 75–86.
4. Golic J., Morgari G. On the resynchronization attack // Fast Software Encryption (FSE'03): Proc. — Berlin: Springer-Verlag, 2003. — P. 100–110.
5. Алексеев Е.К. О некоторых мерах нелинейности булевых функций // Прикл. дискрет. математика. — 2011. — № 2(12). — С. 5–16.
6. Алексеев Е.К. Об атаке на фильтрующий генератор с функцией усложнения, близкой к алгебраически вырожденной // Материалы VI Междунар. научн. конф. по проблемам безопасности и противодействия терроризму, М., 11–12 нояб. 2010 г. — М.: МЦНМО, 2011. — Т. 2. — С. 114–123.
7. Levin L.A. Randomness and non-determinism // J. Symbolic Logic. — 1993. — **58**, N 3. — P. 1102–1103.
8. Testing Reed-Muller codes / N. Alon, T. Kaufman, M. Krivelevich et al // IEEE Trans. Inform. Theory. — 2005. — **51**(11). — P. 4032–4039.
9. Optimal testing of Reed-Muller codes / A. Bhattacharyya, S. Kopparty, G. Schoenebeck at al. // Proc. of the 51st Ann. IEEE Symp. on Fundations of Comput. Science, Las Vegas, Oct. 23–26, 2010. — P. 488–497.

Поступила 23.02.2012