ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 28 МАТЕМАТИКИ Выпуск 2
2021

Д. С. Богданов[1], **В. О. Миронкин**[2] (Москва, Национальный исследовательский ядерный университет «МИФИ», [2] Национальный исследовательский университет «Высшая школа экономики»). **Об эффективности алгоритмов формирования равновероятных последовательностей произвольного модуля на основе схемы независимых равновероятных испытаний Бернулли.**

УДК 519.212.2+004.032.2

Резюме: Рассматривается один из подходов к оценке эффективности алгоритмов преобразования данных, представляющих реализацию схемы независимых равновероятных испытаний Бернулли, в последовательность равновероятных элементов по модулю $q > 2$. Вычислена оценка сверху эффективности одного класса таких алгоритмов.

Ключевые слова: Метод отбраковки, кодирование, разложение по степеням.

Введение. Результаты исследований (см., например, [2]), связанных с разработкой подходов к преобразованию данных, представляющих собой реализацию схемы независимых равновероятных испытаний Бернулли (далее — исходной последовательности), в последовательность равновероятных элементов по модулю $q > 2$ (далее — результирующей последовательности), находят свое применение в ряде практических приложений защиты информации (в частности, при генерации PIN-кодов, паролей и другой аутентифицирующей информации).

Наиболее широко используемым и теоретически обоснованным алгоритмом, позволяющим реализовать указанное преобразование, является так называемый «метод полной отбраковки» [1]. Указанный алгоритм при определенных условиях гарантирует равновероятное распределения знаков выходной последовательности. Вместе с этим, его выполнение в ряде случаев влечет за собой большой «перерасход» бит исходной последовательности (иной раз двукратный). По этой причине возникает естественная задача, заключающаяся в построении более «экономных» алгоритмов преобразования бит исходной последовательности в элементы результирующей.

1. Метод полной отбраковки. Для произвольных $n \in \mathbb{N}$ и $\overline{x} \in V_n$ через $r_n(\overline{x})$ обозначим число, равное $2^0 \cdot x_1 + 2^1 \cdot x_2 + \dots + 2^{n-1} \cdot x_n$. Согласно [1] для формирования одного знака $y \in \mathbb{Z}_q$, $q > 2$, необходимо предварительно определить $n = \min\{t \in \mathbb{N} | 2^t \geqslant q\}$ — количество бит исходной последовательности.

Алгоритм 1. Метод полной отбраковки
Вход: q, n;
1: **Цикл**
2: $\overline{x} \leftarrow$ следующие n бит исходной двоичной последовательности;
3: $y \leftarrow r_n(\overline{x})$;
4: **Если** $y < q$ **то**
5: **Выход:** y
конец цикла

Замечание 1. Если $q \approx 2^{n-1} + 1$, то с вероятностью $\approx \frac{1}{2}$ на шаге 4 алгоритма значение y будет больше либо равно q, т. е. алгоритм не выйдет из цикла и продолжит работу.

© Редакция журнала «ОПиПМ», 2021 г.

Заметим, что если на шаге 4 алгоритма не выполнено сравнение $y < q$, то все текущие n бит исходной последовательности отбрасываются, что, очевидно, является недостатком данного алгоритма. Таким образом, становится актуальным вопрос, связанный с оценкой возможности использования некоторой доли отбракованных (пусть и не всех) бит исходной последовательности для формирования знаков результирующей последовательности.

Таким образом, в случае $y \in \{q, \ldots, 2^n - 1\}$ необходимо построить некоторое «решающее правило», по которому можно определить, какие биты последовательности $\overline{x}$ можно «сохранить» (т. е. использовать для формирования строки $\overline{x}$ в следующем цикле на шаге 1). К такому решающему правилу предъявляются следующие естественные требования:

1. **Однозначность.** По значению y однозначно определяются сохраняемые биты последовательности $\overline{x}$.
2. **Равновероятность и независимость.** Если согласно решающему правилу из последовательности $\overline{x}$ были сохранены некоторые t бит, то значения этих бит равновероятны и независимы.

Разобьем множество элементов $\{q, \ldots, 2^n - 1\}$ на не пересекающиеся классы $A_1, A_2, \ldots, A_i$, $i \in \mathbb{N}$, такие, что в одном классе A_j, $j = 1, \ldots, i$, содержатся все элементы, для которых при условии $y \in A_j$ по решающему правилу сохраняются одни и те же t_j бит исходной последовательности.

Алгоритм 2. Общий вид решающего правила
Вход: $y \in \{q, \ldots, 2^n - 1\}$;
Если $y \in A_1$ **то**
Сохраняются следующие t_1 бит последовательности $\overline{x}$: ...
Если $y \in A_2$ **то**
Сохраняются следующие t_2 бит последовательности $\overline{x}$: ...
$\vdots$
Если $y \in A_i$ **то**
Сохраняются следующие t_i бит последовательности $\overline{x}$: ...

2. Определение «эффективности решающего правила»

Определение 1. Пусть дано решающее правило A, заданное в виде алгоритма . *Эффективностью решающего правила* A назовем величину

$$Eff(A) = \sum_{j=1}^{i} |A_j| \cdot t_j.$$

Замечание 2. Величина $\frac{Eff(A)}{2^n - 1 - q}$ совпадает со средним числа бит исходной последовательности, сохраняемых решающим правилом A.

Определим среди всех решающих правил такое правило, которое при выполнения условия $y \geqslant q$ в среднем сохраняет наибольшее число бит исходной последовательности.

Определение 2. Решающее правило A называется *оптимальным*, если

$$Eff(A) = \max\{Eff(A') | A' \in \mathcal{A}\},$$

где $\mathcal{A}$ — множество всех возможные решающих правила со свойствами 1), 2).

Утверждение 1. *Пусть согласно решающему правилу A при попадании y в A_j, $j = 1, \ldots, i$, где $i\mathbb{N}$, сохраняется t_j бит. Тогда $|A_j| = k \cdot 2^{t_j}$, где $k \in \mathbb{N}$.*

Следствие 1. *Пусть $|A_j| \leqslant 2^t$. Тогда при попадании y в A_j может сохраняется t_j бит, где $t_j \leqslant t$, причем равенство достигается тогда и только тогда, когда $|A_j| = 2^t$.*

Теорема 1. *Пусть* $2^n - 1 - q = 2^{d_1} + \dots + 2^{d_k}$, *где* $0 \leqslant d_1 < d_2 < \dots < d_k$. *Тогда для любого решающего правила* A *со свойствами однозначности, равновероятности и независимости справедливо неравенство*

$$Eff(A) \leqslant \sum_{i=1}^{k} 2^{d_i} \cdot d_i.$$

Теорема 2. *Пусть* $2^n - 1 - q = 2^{d_1} + \dots + 2^{d_k}$, *где* $0 \leqslant d_1 < d_2 < \dots < d_k$. *Тогда для оптимального решающего правила* A *со свойствами однозначности, равновероятности и независимости выполняется равенство*

$$Eff(A) = \sum_{i=1}^{k} 2^{d_i} \cdot d_i.$$

Предъявим решающее правило A со свойствами однозначности, равновероятности и независимости, для которого $Eff(A) = \sum_{i=1}^{k} 2^{d_i} \cdot d_i$.

Итак, пусть q — модуль, по которому вырабатываются знаки результирующей последовательности, $n \in \mathbb{N}:\ 2^{n-1} < q \leqslant 2^n$. Пусть при этом $2^n - 1 - q = 2^{d_1} + 2^{d_2} \dots + 2^{d_k}$, где $0 \leqslant d_1 < d_2 < \dots < d_k \leqslant n-2$. Обозначим через $p_0 = \frac{q}{2^n}$, $p_1 = \frac{2^{d_1}}{2^n}$, $\dots$, $p_k = \frac{2^{d_k}}{2^n}$.

Алгоритм 3. Жадный метод отбраковки

Вход: $q, n, 2^{d_1}, 2^{d_2}, \dots, 2^{d_k}$;

1: $\overline{x} \leftarrow \emptyset$;

2: **Цикл**

3: Дополняем $\overline{x}$ битами исходной двоичной последовательности до строки длины n;

4: $y \leftarrow r_n(\overline{x})$;

5: **Если** $y \in \{0, \dots, q-1\}$ **то**

6: **Выход:** y

7: **Если** $y \in \{q, \dots, q + 2^{d_k} - 1\}$ **то**

8: $\overline{x} \leftarrow \mathrm{LSB}_{d_k}(\overline{x})$

9: **Если** $y \in \{q + 2^{d_k}, \dots, q + 2^{d_k} + 2^{d_{k-1}} - 1\}$

10: $\overline{x} \leftarrow \mathrm{LSB}_{d_{k-1}}(\overline{x})$

11: $\vdots$

12:: **Если** $y \in \{q + 2^{d_k} + \ldots + 2^{d_2}, \dots, q + 2^{d_k} + \ldots + 2^{d_1} - 1\}$ **то**

13: $\overline{x} \leftarrow \mathrm{LSB}_{d_1}(\overline{x})$

конец цикла

СПИСОК ЛИТЕРАТУРЫ

1. *Мартышенко С. Н.* Компьютерный анализ данных: Учебное пособие. — ВГУЭС, 2010, 80 с.
2. *Миронкин В. О.* Об алгоритме формирования равновероятных последовательностей произвольного модуля на основе схемы независимых равновероятных испытаний Бернулли. — Обозрение прикл. и промышл. матем., 2021, т. 28, в. 1, с. ??-??.

UDC 519.212.2+004.032.2

Bogdanov D. S. [1], ***Mironkin V. O.*** [2] ([1] National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Russia, [2] National Research University Higher School of Economics,Russia). **On the efficiency of algorithms for forming equiprobable sequences of an arbitrary module based on the scheme of independent equiprobable Bernoulli tests**

Abstract: One of the approaches to assessing the efficiency of data transformation algorithms, representing the implementation of the scheme of independent equiprobable Bernoulli trials into a sequence of equiprobable elements according to to the module $q > 2$, is considered. An upper bound for the efficiency of one class of such algorithms is calculated.

Keywords: Culling method, encoding, power expansion.