В. О. Миронкин, В. Г. Михайлов, О множестве образов k-кратной итерации равновероятного случайного отображения, *Матем. вопр. криптогр.*, 2018, том 9, выпуск 3, 99–108

DOI: https://doi.org/10.4213/mvk264

МАТЕМАТИЧЕСКИЕ ВОПРОСЫ КРИПТОГРАФИИ
2018 Т. 9 № 3 С. 99–108

УДК 519.212.2+519.719.2 DOI https://doi.org/10.4213/mvk264

О множестве образов k-кратной итерации равновероятного случайного отображения

В. О. Миронкин [1], **В. Г. Михайлов** [2]

[1] *Национальный исследовательский университет «Высшая школа экономики», Москва*

[2] *Математический институт им. В. А. Стеклова РАН, Москва*

Получено 11.V.2017
Переработанный вариант 05.VI.2018

Аннотация. Изучаются свойства графа k-кратной итерации равновероятного случайного отображения $f\colon \{1,\dots,n\} \to \{1,\dots,n\}$. Получены рекуррентные формулы для вероятностей принадлежности вершины множеству $f^k(\{1,\dots,n\})$ и множеству висячих вершин в графе отображения f^k.

Ключевые слова: равновероятное случайное отображение, степень отображения, граф отображения, образ, прообраз, висячая вершина

On the sets of images of k-fold iteration of uniform random mapping

V. O. Mironkin [1], **V. G. Mikhailov** [2]

[1] *National Research University Higher School of Economics, Moscow*
[2] *Steklov Mathematical Institute of RAS, Moscow*

Abstract. The properties of the graph of k-fold iteration of uniform random mapping $f : \{1,\dots,n\} \to \{1,\dots,n\}$ are studied. Some recurrence formulas for the probabilities for a vertex to belong to the set of images $f^k(\{1,\dots,n\})$ and to the set of the initial vertices in the graph of f^k are obtained.

Key words: uniform random mapping, iteration of random mapping, graph of a mapping, image, pre-image, initial vertex

Citation: *Mathematical Aspects of Cryptography*, 2018, v. 9, № 3, pp. 99–108 (Russian)

1. Введение

Исследования моделей отображений, построенных на основе равновероятных случайных отображений [1, 4, 14, 15], представляют особый интерес для практических приложений современной криптографии [8 – 10, 16 – 18]. Так, одной из наиболее важных задач, возникающих при описании вероятностных свойств итерационных алгоритмов выработки производных ключей на основе некоторой долговременной информации, является оценка степени сжатия исходного ключевого множества через фиксированные числа итераций. В ряде случаев значение указанной характеристики позволяет оценить стойкость криптографического примитива относительно метода тотального опробования [12].

Наряду с оценкой мощности ключевого множества практический интерес представляет и обратная задача, заключающаяся в определении множества ключей, выработанных к моменту наблюдения. Такая информация может быть использована при компрометации текущего ключа.

Настоящая статья посвящена изучению k-кратных итераций равновероятного случайного отображения [3, 11, 19, 20] и продолжает цикл исследований, начатых в [2, 6, 7]. Объектом исследования является образ множества $\{1, \ldots, n\}$ при действии k-кратной итерации случайного отображения, распределенного равномерно на множестве $\{1, \ldots, n\}$.

В работе среди прочих используются следующие определения для характеристик графа отображения (в определениях отображение f считается детерминированным).

Определение 1. Вершина $x_0 \in S$ называется *циклической вершиной* графа G_f отображения f, если существует такое $b \geqslant 1$, что

$$f^b(x_0) = x_0.$$

Множество циклических вершин графа G_f обозначим $C(G_f)$.

Высотой $\alpha_f(x_0)$ вершины $x_0 \in S$ в графе G_f называется расстояние от этой вершины до ближайшей циклической вершины:

$$\alpha_f(x_0) = \min\{m \geqslant 0 \colon f^m(x_0) \in C(G_f)\}.$$

Через $\beta_f(x_0)$ обозначим длину цикла компоненты графа G_f, содержащей вершину x_0.

2. Формулировки результатов

Следуя [2], рассмотрим множество $S = \{1, \ldots, n\}$, $n \geqslant 2$, и вероятностное пространство $(\Omega, \mathcal{F}, \mathbf{P})$, где элементарные исходы образуют множество

$$\Omega = \{f \colon S \to S\}$$

всех n^n отображений S в себя, алгебра событий $\mathcal{F}$ состоит из всех подмножеств Ω, а вероятностная мера $\mathbf{P}$ является равномерной:

$$\mathbf{P}(f) = \frac{1}{n^n} \quad \forall\, f \in \Omega.$$

Положим

$$S(r) = \{1, \ldots, r\}, \quad \text{где} \quad 1 \leqslant r < n.$$

Для произвольного $y \in S(r)$ рассмотрим следующее событие:

$$F_y^{(k)}(r) = \big\{y, f(y), \ldots, f^{k-1}(y) \in S(r) \text{ различны; } f^k(y) = r+1\big\}.$$

Для произвольного $m \in \overline{1, r}$ определим величины

$$B_{r,m}^{(k)} = \sum_{1 \leqslant y_1 < \ldots < y_m \leqslant r} \mathbf{P}\Big\{ \bigcap_{i=1}^{m} F_{y_i}^{(k)}(r) \Big\} = C_r^m P_{r,m}^{(k)}, \tag{1}$$

где

$$P_{r,m}^{(k)} = \mathbf{P}\Big\{ \bigcap_{i=1}^{m} F_i^{(k)}(r) \Big\}.$$

При этом событие

$$\Big\{ \bigcap_{i=1}^{m} F_i^{(k)}(r) \Big\},$$

указанное в правой части (1), означает, что в течение первых $k-1$ итераций отображение f не позволяет зацикливаться образам вершин $1, \ldots, m$ и не выводит их из множества $S(r)$, а на k-м шаге отображение f переводит их в одну и ту же вершину $r+1$. Из (1) легко вывести, что

$$P_{r,m}^{(1)} = \frac{1}{n^m}, \quad B_{r,m}^{(1)} = \frac{C_r^m}{n^m}. \tag{2}$$

Следующее утверждение удобно привести непосредственно перед формулировкой основного результата статьи.

Лемма. *Пусть случайное отображение* $f\colon S \to S$ *имеет равномерное распределение на* Ω. *Тогда величины* $P_{r,m}^{(k)}$, *определенные формулами* (1) *и* (2), *при* $m = 1, \ldots, r$ *удовлетворяют соотношениям*

$$P_{r,m}^{(k)} = \sum_{s=1}^{m} a_{r,m,s}\, P_{r-m,s}^{(k-1)}, \qquad k \geqslant 2, \tag{3}$$

где числа $a_{r,m,s}$ *определяются равенствами*

$$a_{r,m,s} = C_{r-m}^{s} \Big(\frac{s}{n}\Big)^{m} \sum_{t=0}^{s} C_s^t\, (-1)^t \Big(1 - \frac{t}{s}\Big)^{m}. \tag{4}$$

Через $f^k(S)$, $k \geqslant 1$, обозначим образ множества S при отображении f^k.

Теорема. *Пусть случайное отображение* $f\colon S \to S$ *имеет равномерное распределение на* Ω. *Тогда при любых* $k \geqslant 1$, $x_0 \in S$ *справедливо равенство*

$$\mathbf{P}\big\{x_0 \in f^k(S)\big\} = \sum_{l=1}^{n} \frac{(n)_l}{n^{l+1}} + \sum_{t=1}^{n-1} \sum_{l=1}^{n-t} \frac{(n-1)_{r+l-1}}{n^{r+l}} \sum_{m=1}^{n-l-t} (-1)^m B_{n-l-t,m}^{(k)}, \tag{5}$$

где величины $B_{n-l-t,m}^{(k)}$ *удовлетворяют равенству* (2).

3. Доказательства

Доказательство леммы. Пусть $m = 1$. Тогда согласно (1) имеет место равенство

$$P_{r,1}^{(k)} = \mathbf{P}\big\{F_1^{(k)}(r)\big\} = \frac{(r-1)\,(r-2)\cdots(r-k+1)\cdot 1}{n^k} = \frac{(r-1)_{k-1}}{n^k},$$

где

$$(q)_k = q\,(q-1)\cdots(q-k+1)$$

— k-я факториальная степень числа q, что соответствует (2) и (3).

Пусть $m \geqslant 2$. Для случайного равновероятного отображения $f\colon S \to S$ совместное распределение значений $f(1), \ldots, f(m)$ соответствует равновероятной схеме независимого размещения m частиц по n ячейкам, занумерованным числами $1, \ldots, m, m+1, \ldots, n$.

Через $G_m(r)$ обозначим событие, состоящее в том, что в результате размещения все m частиц попали в ячейки с номерами из множества $S(r)\backslash\{1,\ldots,m\} = \{m+1,\ldots,r\}$, а через $\mu(m,n)$ — случайную величину, равную числу занятых ячеек.

Для значений $f(1),\ldots,f(m)$ рассмотрим полную группу событий $H_s = \{\mu(m,n)=s\}$, $s=1,2,\ldots,m$, соответствующих тому, что после применения отображения f к вершинам $1,\ldots,m$ соответствующие m траекторий в графе G_f склеились ровно в s траекторий (см. рис. 1).

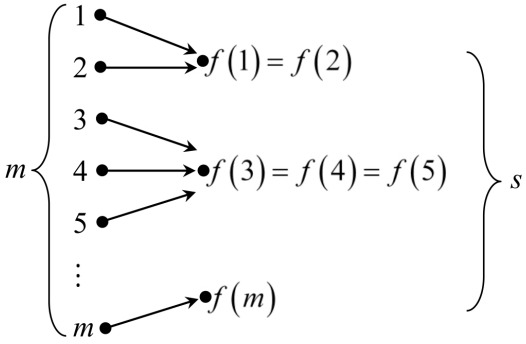

Рис. 1. Пример склеивания траекторий в графе G_f

Таким образом, по формуле полной вероятности получаем рекуррентное соотношение:

$$P_{r,m}^{(k)} = \sum_{s=1}^{m} \mathbf{P}\big\{G_m(r) \cap \{\mu(m,n)=s\}\big\}\, P_{r-m,s}^{(k-1)}. \tag{6}$$

При каждом фиксированном $s \in \{1,\ldots,m\}$ представим вероятность, стоящую под знаком суммирования в (6), в следующем виде:

$$\mathbf{P}\big\{G_m(r) \cap \{\mu(m,n)=s\}\,\big\} = \mathbf{P}\{\mu(m,n)=s \,|\, G_m(r)\}\,\mathbf{P}\{G_m(r)\}.$$

При этом

$$\mathbf{P}\{G_m(r)\} = \Big(\frac{r-m}{n}\Big)^m, \tag{7}$$

$$\begin{aligned}\mathbf{P}\{\mu(m,n)=s \,|\, G_m(r)\} &= \mathbf{P}\{\mu(m,r-m)=s\} = \\ &= \mathbf{P}\{\mu(m,r')=s\} = \mathbf{P}\{\mu_0(m,r')=r'-s\},\end{aligned} \tag{8}$$

где $r' = r-m$, а $\mu_0(m,r') = r' - \mu(m,r')$ — число пустых ячеек при равновероятном размещении m частиц по r' ячейкам.

Согласно формулам для распределения величины $\mu_0(m, r')$ (см. [5], формулы (1) и (2) на с. 10)

$$\mathbf{P}\{\mu_0(m,r')=r'-s\} = C_{r'}^{s}\,\Big(\frac{s}{r'}\Big)^m \sum_{t=0}^{s} C_s^t\,(-1)^t\Big(1-\frac{t}{s}\Big)^m. \tag{9}$$

Из (7) — (9) вытекает формула

$$\mathbf{P}\big\{G_m(r)\cap\{\mu(m,n)=s\}\big\} = C_{r-m}^{s}\Big(\frac{s}{n}\Big)^m \sum_{t=0}^{s} C_s^t\,(-1)^t\Big(1-\frac{t}{s}\Big)^m.$$

Из нее и (6) следуют (3) и (4). Лемма доказана. □

Доказательство теоремы. Для произвольной вершины $x_0 \in S$ рассмотрим событие

$$E_{t,l}(x_0) = \big\{\alpha_f(x_0)=t,\ \beta_f(x_0)=l\big\},$$

которое может быть представлено в виде объединения попарно несовместных событий:

$$E_{t,l}(x_0) = \bigcup_{T_{t,l}} \big\{f(x_0)=x_1,\ldots,f^{t+l}(x_0)=x_{t+l}\big\},$$

где

$$T_{t,l} = \{(x_1,\ldots,x_{t+l}):\\ x_1,\ldots,x_{t+l}\in S\setminus\{x_0\},\ x_{t+l}=x_t,\ \big|\{x_1,\ldots,x_{t+l}\}\big|=t+l-1\}.$$

Здесь и далее t и l указывают соответственно высоту вершины x_0 и длину цикла, на подходе к которому лежит x_0.

Очевидно, что

$$\big\{x_0\in f^k(S)\big\} = \big\{x_0\in C(G_{f^k})\big\}\cup\bigcup_{t=1}^{n-1}\bigcup_{l=1}^{n-t}\big\{\big\{x_0\in f^k(S)\big\}\cap E_{t,l}(x_0)\big\}, \tag{10}$$

где, напомним, $C(G_{f^k})$ — множество циклических вершин графа G_{f^k}.

В силу несовместности событий, указанных в правой части (10), и равенства $C(G_{f^k}) = C(G_f)$ получаем

$$\mathbf{P}\big\{x_0\in f^k(S)\big\} = \mathbf{P}\big\{x_0\in C(G_f)\big\} + \sum_{t=1}^{n-1}\sum_{l=1}^{n-t}\mathbf{P}\big\{\big\{x_0\in f^k(S)\big\}\cap E_{t,l}(x_0)\big\}. \tag{11}$$

Для первого слагаемого в (11) выполнено равенство (см., например, [15])

$$\mathbf{P}\big\{x_0\in C(G_f)\big\} = \sum_{l=1}^{n}\frac{(n)_l}{n^{l+1}}. \tag{12}$$

Теперь вычислим вероятности, стоящие под знаками суммирования в правой части (11). При фиксированных значениях $t\in\{1,\dots,n-1\}, l\in\{1,\dots,n-t\}$ имеет место соотношение

$$\begin{aligned}&\{x_0\in f^k(S)\}\cap E_{t,l}(x_0)=\\&=\bigcup_{T_{t,l}}\bigcup_{y\in S'}\{\widetilde{F}^{(k)}_{x_0,y}(S');\ f(x_0)=x_1,\dots,f^{t+l}(x_0)=x_{t+l}\},\end{aligned}\tag{13}$$

где $S'=S\setminus\{x_0,x_1,\dots,x_{t+l-1}\}$ и для произвольного $M\subseteq S$

$$\widetilde{F}^{(k)}_{x_0,y}(M)=\{y,f(y),\dots,f^{k-1}(y)\in M \text{ различны};\ f^k(y)=x_0\}.$$

Из несовместности событий $\{f(x_0)=x_1,\dots,f^{t+l}(x_0)=x_{t+l}\}$, входящих в правую часть равенства (13), получаем, что

$$\begin{aligned}&\mathbf{P}\{\{x_0\in f^k(S)\}\cap E_{t,l}(x_0)\}=\\&=\sum_{T_{t,l}}\mathbf{P}\Big\{\bigcup_{y\in S'}\{\widetilde{F}^{(k)}_{x_0,y}(S');\ f(x_0)=x_1,\dots,f^{t+l}(x_0)=x_{t+l}\}\Big\}.\end{aligned}\tag{14}$$

Заметим, что все слагаемые суммы в (14) одинаковы по величине. Рассмотрим слагаемое, в котором

$$\begin{aligned}&(x_1,\dots,x_{t+l})=v=\\&=(n-t-l+2,n-t-l+3,\dots,n-l+1,\dots,n,n-l+1)\in T_{t,l}.\end{aligned}$$

Для него $S'=\{1,\dots,r\}$, $r=n-t-l$, и $x_0=r+1$. Умножим это слагаемое на общее число слагаемых $|T_{t,l}|=(n-1)_{t+l-1}$ и воспользуемся независимостью в совокупности случайных величин $f(x)$, $x\in S$:

$$\begin{aligned}&\mathbf{P}\{\{x_0\in f^k(S)\}\cap E_{t,l}(x_0)\}=\\&=(n-1)_{t+l-1}\,\mathbf{P}\Big\{\bigcup_{y\in S'}\widetilde{F}^{(k)}_{x_0,y}(S')\Big\}\,\mathbf{P}\{(f(x_0),\dots,f^{t+l}(x_0))=v\}=\\&=\frac{(n-1)_{t+l-1}}{n^{t+l}}\,\mathbf{P}\Big\{\bigcup_{y\in S'}\widetilde{F}^{(k)}_{x_0,y}(S')\Big\}=\\&=\frac{(n-1)_{t+l-1}}{n^{t+l}}\,\mathbf{P}\Big\{\bigcup_{y\in S(r)}F^{(k)}_{y}(r)\Big\}.\end{aligned}\tag{15}$$

По формуле включения-исключения (см, например, [13])

$$\mathbf{P}\Big\{\bigcup_{y\in S(r)} F_y^{(k)}(r)\Big\} = \sum_{m=1}^{r}(-1)^m B_{r,m}^{(k)}, \tag{16}$$

а величины $B_{r,m}^{(k)}$ определяются соотношением (1). Таким образом, из (15), (16) следует, что

$$\mathbf{P}\Big\{\big\{x_0 \in f^k(S)\big\} \cap E_{t,l}(x_0)\Big\} = \frac{(n-1)_{t+l-1}}{n^{t+l}} \sum_{m=1}^{r}(-1)^m B_{r,m}^{(k)}. \tag{17}$$

Подставив (17) и (12) в (11), получим искомое выражение (5).

Теорема доказана. □

Замечание 1. Теорема позволяет вычислить среднее значение мощности множества S при действии отображения f^k. Так как

$$\big|f^k(S)\big| = \sum_{x_0\in S} I\big\{x_0 \in f^k(S)\big\},$$

то в силу равноправия всех $x_0 \in S$

$$\mathbf{E}\big|f^k(S)\big| = \mathbf{E}\sum_{x_0\in S} I\big\{x_0 \in f^k(S)\big\} = n\,\mathbf{P}\big\{x_0 \in f^k(S)\big\}.$$

Замечание 2. В случае когда величина k превышает максимальную длину подхода в графе G_{f^k}, что заведомо выполняется при $k \geqslant n-1$, имеет место равенство $f^k(S) = C(G_f)$.

Определение 2. Вершина $x_0 \in S$ в графе G_f называется *висячей вершиной*, если не существует такого $y \in S$, что $f(y) = x_0$.

Из определения 2 следует, что множество висячих вершин графа G_f совпадает с множеством вершин, не имеющих прообразы.

Через T_{f^k} обозначим множество висячих вершин в графе G_{f^k}, $k \in \mathbb{N}$. Множество T_f исследовалось, например, в [14].

Рассмотрим случай $k \geqslant 2$. Найдем вероятность попадания случайной вершины графа G_{f^k} в множество T_{f^k}.

Заметим, что для любого отображения $f\colon S \to S$

$$S = f^k(S) \cup T_{f^k}, \quad f^k(S) \cap T_{f^k} = \varnothing.$$

Поэтому для произвольного $x_0 \in S$ справедливо соотношение

$$\mathbf{P}\left\{x_0 \in T_{f^k}\right\} = 1 - \mathbf{P}\left\{x_0 \in f^k(S)\right\},$$

позволяющее применить теорему для вычисления $\mathbf{P}\{x_0 \in T_{f^k}\}$.

Следствие. *Пусть случайное отображение $f\colon S \to S$ имеет равномерное распределение на Ω. Тогда при любых $k \geqslant 1$, $x_0 \in S$ справедливо равенство*

$$\mathbf{P}\{x_0 \in T_{f^k}\} = 1 - \sum_{l=1}^{n} \frac{(n)_l}{n^{l+1}} - \\ - \sum_{t=1}^{n-1} \sum_{l=1}^{n-t} \frac{(n-1)_{r+l-1}}{n^{r+l}} \sum_{m=1}^{n-l-t} (-1)^m B^{(k)}_{n-l-t,m},$$

где величины $B^{(k)}_{n-l-t,m}$ удовлетворяют равенству (2).

Замечание 3. Аналогично замечанию 1 получаем, что

$$\mathbf{E}\left|T_{f^k}\right| = n\,\mathbf{P}\{x_0 \in T_{f^k}\} = n - n\,\mathbf{P}\{x_0 \in f^k(S)\}.$$

Замечание 4. При $k \geqslant n-1$ множеству T_{f^k} принадлежат все вершины графа G_f, лежащие на подходах к циклам.

Замечание 5. Полученные в настоящей статье результаты могут найти свое применение не только при описании свойств криптографических примитивов, имеющих итерационную структуру (хэш-функции, генераторы ключей, датчики псевдослучайных последовательностей), но и при решении задачи анализа качества псевдослучайных последовательностей, вырабатываемых на основе некоторого итерационного преобразования.

Авторы признательны А. М. Зубкову за интерес к работе и полезные замечания.

Список литературы

[1] Зубков А. М., “Вычисление распределения характеристик числа компонент и циклических точек случайного отображения”, *Математические вопросы криптографии*, № 2 (2010), 5–18.

[2] Зубков А.М., Миронкин В. О., “Распределение длины отрезка апериодичности в графе k-кратной итерации случайного равновероятного отображения”, *Математические вопросы криптографии*, **8**:4 (2017), 63–74.

[3] Зубков А. М., Серов А. А., “Совокупность образов подмножества конечного множества при итерациях случайных отображений”, *Дискретная математика*, **26**:4 (2014), 43–50.

[4] Колчин В. Ф., *Случайные отображения*, М.: Наука, 1984, 208 с.

[5] Колчин В. Ф., Севастьянов Б. А., Чистяков В. П., *Случайные размещения*, М.: Наука, 1976, 224 с.

[6] Миронкин В. О., “Исследование свойств и характеристик степени случайного отображения”, *Обозрение прикл. и промышл. матем.*, **21**:1 (2014), 70–73.

[7] Миронкин В. О., “Об особенностях строения графа степени случайного отображения”, *Обозрение прикл. и промышл. матем.*, **23**:1 (2016), 57–62.

[8] Миронкин В. О., “О некоторых вероятностных характеристиках алгоритма выработки ключа «CRYPTOPRO KEY MESHING»”, *Проблемы информационной безопасности. Компьютерные системы*, № 4 (2015), 140–146.

[9] Михайлов В. Г., “Исследование комбинаторно-вероятностной модели автоматов из регистров с неравномерным движением”, *Труды по дискретной математике* (2002), 139–149.

[10] Михайлов В. Г., “Исследование числа циклических точек автомата из регистров с неравномерным движением”, *Труды по дискретной математике* (2002), 167–172.

[11] Пильщиков Д. В., “Асимптотическое поведение мощности полного прообраза случайного множества при итерациях отображений конечного множества”, *Математические вопросы криптографии*, **8**:1 (2017), 95–106.

[12] Погорелов Б. А., Сачков В. Н., *Словарь криптографических терминов*, М.: МЦНМО, 2006, 94 с.

[13] Сачков В. Н., *Вероятностные методы в комбинаторном анализе*, М.: Наука, 1978, 288 с.

[14] Flajolet P., Odlyzko A., “Random mapping statistics”, Lect. Notes Comput. Sci., **434**, 1989, 329–354.

[15] Harris B., “Probability distributions related to random mapping”, *Ann. Math. Statist.*, **31**:4 (1960), 1045–1062.

[16] Hellman M. E., “A cryptanalytic time–memory trade-off”, *IEEE Trans. Inf. Theory* (1980), 401–406.

[17] Hong J., Ma D., “Success probability of the Hellman trade-off”, *Inf. Process. Lett.*, **109**:7 (2009), 347–351.

[18] Oechslin P., “Making a faster cryptanalytic time-memory trade-off”, *Lect. Notes Comput. Sci.*, **2729** (2003), 617–630.

[19] Pilshchikov D. V., “Estimation of the characteristics of time-memory-data tradeoff methods via generating functions of the number of particles and the total number of particles in the Galton-Watson process”, *Математические вопросы криптографии*, **5**:2 (2014), 103–108.

[20] Pilshchikov D. V., “On the limiting mean values in probabilistic models of time-memory-data tradeoff methods”, *Математические вопросы криптографии*, **6**:2 (2015), 59–65.