Math-Net.Ru

Общероссийский математический портал

Д. В. Пильщиков, Исследование сложности метода радужных таблиц с маркерами цепочек, *Матем. вопр. криптогр.*, 2017, том 8, выпуск 4, 99–116

DOI: <https://doi.org/10.4213/mvk241>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.162.27.143

26 апреля 2023 г., 16:05:59

Исследование сложности метода радужных таблиц с маркерами цепочек

Д. В. Пильщиков

Лаборатории ТВП, Москва

Получено 11.V.2017

Аннотация. Вводится вероятностная модель оперативного этапа метода радужных таблиц с маркерами цепочек и на ее основе рассчитывается средняя временная сложность полной обработки одной таблицы. Выводятся двусторонние оценки этой величины и предлагается вычислительный подход к построению оптимального маркера. Используемая вероятностная модель в отличие от других моделей позволяет учитывать влияние дисперсии числа прообразов случайно и равновероятно выбранного элемента конечного множества относительно однонаправленной функции.

Ключевые слова: балансировка времени-памяти-данных, радужные таблицы, вероятностные модели, ветвящиеся процессы

Complexity analysis of the method of rainbow tables with fingerprints

D. V. Pilshchikov

TVP Laboratories, Moscow

Abstract. The probabilistic model of the operative stage of the rainbow tables with fingerprints method is introduced. It is used to compute the mean time complexity of the complete processing of one table. Two-sided estimates of this complexity are obtained and computational approach to the construction of the optimal fingerprint is suggested. Our probabilistic model as opposed to other models permits to account for the impact of the variance of the preimages number of random uniformly chosen element of a finite set with respect to the one-way function.

Key words: time-memory-data tradeoff, rainbow tables, probabilistic models, branching processes

Citation: *Mathematical Aspects of Cryptography*, 2017, v. 8, № 4, pp. 99–116 (Russian)

© Академия криптографии Российской Федерации, 2017 г.

Введение

Ряд задач теоретической и прикладной криптографии [1–7] связан с задачей обращения однонаправленной функции

$$G: X \rightarrow X, \quad |X| = N,$$

в одной из точек набора $\mathcal{D} = \{a_1, \dots, a_D\}$, $a_i \in X$, $i \in \overline{1, D}$.

Для ее решения широкое распространение получили переборные алгоритмы типа балансировки времени–памяти–данных, или TMD-балансировки [1–3, 8–11].

Алгоритмы данного класса состоят из двух этапов. На первом (предварительном) этапе проводятся трудоемкие вычисления, результаты которых записываются в таблицах. На втором (оперативном) этапе с использованием таблиц решается задача обращения однонаправленной функции. Таблицы строятся однократно и используются для обращения произвольного числа данных. Временная сложность обращения данных на оперативном этапе значительно меньше временной сложности тотального перебора.

Наиболее эффективными алгоритмами TMD-балансировки признаны балансировки с радужными таблицами [3] (далее RM-метод) и с совершенными радужными таблицами [8] (далее RMP-метод). Известны три модификации этих методов, позволяющие понизить их емкостную и временную сложности. Первая модификация основана на использовании «проверочных» точек [8, 9]. Вторая модификация основана на «редуцировании конечных точек» [2, 10].

В третьей модификации (метод маркеров цепочек) проверочные точки и редуцированные конечные точки используются одновременно. Первое описание метода маркеров цепочек появилось в электронном архиве Cryptology ePrint Archive (Report 2012/683) и позднее было удалено. В 2015 году оно появилось в [11].

Временная сложность указанных выше модификаций RMP-метода с достаточной полнотой исследована в [8, 9, 11–13, 15, 16, 18].

Результаты аналогичных исследований для RM-метода впервые появились в [12]: была введена вероятностная модель, на основе которой рассчитано сокращение средней сложности оперативного этапа RM-метода с одной проверочной точкой, а также найдена оптимальная позиция этой точки. В [13, 14] эти результаты распространены на случай произвольного числа проверочных точек. В [17] рассчитано увеличение средней сложности оперативного этапа RM-метода при редуцировании конечных точек.

В данной работе исследуется временная сложность оперативного этапа RМ-метода с маркером цепочек (далее RММ-метода). Для этого сначала вводится теоретико-вероятностная модель, на основе которой рассчитывается средняя временная сложность полной обработки одной таблицы. Затем выводятся двусторонние оценки этой величины и предлагается вычислительный подход к построению оптимального маркера. Адекватность сделанных вероятностных предположений подтверждается результатами экспериментов.

Использованная в работе теоретико-вероятностная модель, в отличие от моделей работ [13, 14, 17], позволяет учитывать влияние такой характеристики однонаправленной функций G , как дисперсия числа прообразов случайно и равномерно выбранного элемента множества X .

1. Описание метода маркеров цепочек с радужными таблицами

Пусть $\bar{x} = (x_0, \dots, x_{i-1})$ и $\bar{y} = (y_0, \dots, y_{j-1})$, $j \leq i$, — две цепочки элементов множества X . Будем писать $\bar{y} \leq \bar{x}$, если цепочка $\bar{x}$ оканчивается цепочкой $\bar{y}$. Аналогично, будем писать $h' \leq h$, если битовый вектор h' совпадает с концом битового вектора h .

В основе RММ-метода лежит следующее понятие.

О п р е д е л е н и е 1. Отображение, которое ставит в соответствие произвольной конечной цепочке $\bar{x} = (x_0, \dots, x_{k-1})$ элементов множества X битовый вектор $H(\bar{x})$, называется *маркером цепочки $\bar{x}$* , если из условия $\bar{y} \leq \bar{x}$ следует условие $H(\bar{y}) \leq H(\bar{x})$.

Далее будем полагать, что используемые в изучаемом алгоритме маркеры описываются с помощью последовательности отображений $\varphi_i: X \rightarrow V_{n_i}$, $n_i \in \{0, 1, \dots\}$, $i = 0, 1, \dots$, по формулам

$$H(x_0, \dots, x_{k-1}) = \varphi_{k-1}(x_0) \parallel \varphi_{k-2}(x_1) \parallel \dots \parallel \varphi_1(x_{k-2}) \parallel \varphi_0(x_{k-1}),$$

где V_{n_i} — множество битовых строк длины n_i (считается, что пустая строка имеет длину 0). Те значения индекса i , при которых $n_i > 0$, принято называть *позициями n_i -битных проверочных точек*. Множество позиций проверочных точек обозначим символом I_0 .

В общем виде RММ-метод может использовать несколько радужных таблиц и применяться к нескольким элементам данных. В этом случае к каждой таблице и к каждому элементу данных применяется однотипный алгоритм. Опишем его основную идею.

Перед началом алгоритма произвольным образом выбираются множество начальных точек $S \subset X$, $|S| = m$, множество подстановок

$$\{R_i: X \rightarrow X, i = 1, \dots, t\}$$

и маркер $H(\bar{x})$.

Любое из отображений $F^{[i]}(x) = R_i(G(x))$, $i = 1, \dots, t$, далее будем называть F -операцией. Последовательную композицию F -операций вида

$$F^{[j]} \circ F^{[j-1]} \circ \dots \circ F^{[i+1]} \circ F^{[i]}(x), \quad t \geq j \geq i \geq 1,$$

будем обозначать символом $F_{i,j}(x)$. При $i > j$ будем считать, что $F_{i,j}(x) = x$.

Цепочку $(x_{i-1}, x_i, \dots, x_{j-1}, x_j)$ элементов множества X , полученную последовательным применением F -операций к начальной точке $x_{i-1} = x: x_k = F^{[k]}(x_{k-1})$, $k = i, \dots, j$, обозначим символом $\lambda_{i,j}(x)$. При $i > j$ будем считать, что $\lambda_{i,j}(x) = (x)$.

Решение уравнения вида $G(x) = a$ с помощью описываемого алгоритма происходит в два этапа: предварительный и оперативный.

На предварительном этапе для каждой начальной точки $x \in S$ вычисляется цепочка $\lambda_{1,t}(x)$ (*табличная цепочка*). Ее начало x и маркер $H(\lambda_{1,t}(x))$ записываются в таблице T .

На оперативном этапе выполняется t однотипных шагов. На шаге с номером i , $i = 1, 2, \dots, t$, для точки $x_i = R_i(a)$ вычисляется цепочка $\lambda_{i+1,t}(x_i)$ (*основная цепочка*). Маркер $H(\lambda_{i+1,t}(x_i))$ сравнивается с концами маркеров, находящихся в записях таблицы T . Если для маркера h , находящегося в записи (h, x) , выполняется условие $H(\lambda_{i+1,t}(x_i)) \leq h$, то вычисляется цепочка $\lambda_{1,i-1}(x)$ (*дополнительная цепочка*). Для ее последнего элемента $y = F_{1,i-1}(x)$ выполняется проверка равенства $G(y) = a$. Если оно верно, то элемент y является одним из решений рассматриваемого уравнения.

Расчет надежности, временной сложности вычисления табличных и основных цепочек в RMM-методе производится точно так же, как в RM-методе [17]. Основную проблему составляет расчет временной сложности построения дополнительных цепочек.

2. Описание изучаемой величины

Опишем функцию $T(a)$, $a \in X$, которая задает число F -операций и проверок вида $G(y) = a$, выполненных при построении дополнительных цепочек на всех шагах оперативного этапа.

Данную функцию представим в виде суммы $T(a) = \sum_{i=1}^t i T_i(a)$, где $T_i(a)$, $i \in \{1, \dots, t\}$, — число дополнительных цепочек, построенных для основной цепочки на i -м шаге.

Через $D(z, i)$ обозначим множество элементов $y \in X$, для которых цепочка F -операций $\lambda_{1,i}(y)$ заканчивается в точке z .

Пусть $V(i)$ — множество битовых векторов длины $\sum_{j=0}^{t-i} n_j$; через $\overline{D}(h, i)$, $h \in V(i)$, $i = 0, \dots, t$, обозначим множество элементов $x \in X$, для которых выполнено условие $H(\lambda_{i+1,t}(x)) \leq h$.

С учетом введенных обозначений получаем, что если имеет место событие $\{R_i(a) \in \overline{D}(h, i)\}$, то число дополнительных цепочек $T_i(a)$ равно

$$T_i(a) = \sum_{x \in \overline{D}(h, i)} |D(x, i) \cap S|.$$

Таким образом, верна следующая формула:

$$T(a) = \sum_{i=1}^t i \sum_{h \in V(i)} I\{R_i(a) \in \overline{D}(h, i)\} \sum_{x \in \overline{D}(h, i)} |D(x, i) \cap S|.$$

3. Описание теоретико-вероятностной модели

Опишем теоретико-вероятностную модель, в рамках которой вычислим среднее значение величины $T(a)$, которую и будем рассматривать далее в качестве средней временной сложности построения дополнительных цепочек при полной обработке одной таблицы.

Пусть теперь a — случайный равновероятный элемент множества X . Тогда среднее значение T_1 величины $T(a)$ имеет вид

$$\begin{aligned} T_1 = \mathbf{E} T(a) &= \sum_{i=1}^t i \sum_{h \in V(i)} \mathbf{E} I\{R_i(a) \in \overline{D}(h, i)\} \sum_{x \in \overline{D}(h, i)} |D(x, i) \cap S| = \\ &= \sum_{i=1}^t i \sum_{h \in V(i)} \frac{|\overline{D}(h, i)|}{N} \sum_{x \in \overline{D}(h, i)} |D(x, i) \cap S|. \end{aligned}$$

Пусть теперь каждый элемент множества X независимо от других элементов множества X с вероятностью m/N является и с вероятностью $1 - m/N$ не является началом табличной цепочки. Тогда множество S начал цепочек является случайным и для любого фиксированного множества $A \subset X$ выполняется соотношение

$$\mathbf{E} |A \cap S| = \frac{m}{N} |A|.$$

Так как значение T_1 зависит от S , при сделанном вероятностном предположении оно становится случайной величиной. Среднее значение T_2 случайной величины T_1 имеет вид

$$\begin{aligned}
 T_2 = \mathbf{E} T_1 &= \sum_{i=1}^t i \sum_{h \in V(i)} \frac{|\overline{D}(h, i)|}{N} \sum_{x \in \overline{D}(h, i)} \mathbf{E} |D(x, i) \cap S| = \\
 &= \sum_{i=1}^t i \sum_{h \in V(i)} \frac{|\overline{D}(h, i)|}{N} \sum_{x \in \overline{D}(h, i)} \frac{m}{N} |D(x, i)| = \\
 &= \sum_{i=1}^t \frac{mi}{N^2} \sum_{h \in V(i)} \sum_{x \in \overline{D}(h, i)} |D(x, i)| |\overline{D}(h, i)|.
 \end{aligned} \tag{1}$$

Пусть теперь подстановки $R_1, \dots, R_t$ случайны, равновероятны и независимы друг от друга. Тогда величина T_2 становится случайной и ее среднее значение полностью определяется распределением векторов вида $(|D(x, i)|, |\overline{D}(h, i)|)$. Легко заметить, что координаты этих векторов независимы, распределение первой координаты не зависит от x , а распределение второй координаты не зависит от h .

Опишем теоретико-вероятностный подход, в рамках которого смоделируем распределение случайных величин $|D(x, i)|$ и $|\overline{D}(h, i)|$.

Обозначим через $D(x, j, i)$ случайное множество элементов $y \in X$, для которых цепочка F -операций $\lambda_{j,i}(y)$ заканчивается в точке x . Заметим, что величина $|D(x, i)|$ является последним элементом в цепочке

$$|D(x, i+1, i)|, |D(x, i, i)|, |D(x, i-1, i)|, \dots, |D(x, 1, i)|, \tag{2}$$

причем каждый следующий элемент цепочки выражается через предыдущий элемент цепочки с помощью соотношений

$$|D(x, j, i)| = \sum_{y \in D(x, j+1, i)} |D(y, j, j)|, \quad j = i, i-1, \dots, 1, \tag{3}$$

а первая величина в цепочке всегда равна 1:

$$|D(x, i+1, i)| = |\{x\}| = 1.$$

Распределение каждой из случайных величин

$$|D(y, j, j)| = |F_j^{-1}(y)|, \quad y \in X, \quad j \in 1, \dots, t,$$

стоящих в правой части (3), совпадает с распределением случайной величины η , равной мощности полного прообраза случайной равновероятной точки относительно отображения G

$$\mathbf{P}\{|F_j^{-1}(y)| = k\} = \mathbf{P}\{\eta = k\} = \frac{|\{z \mid |G^{-1}(z)| = k\}|}{N}.$$

Пренебрежем зависимостью между случайными величинами вида $|D(y, j, j)|$ в формуле (3): будем считать, что они независимы в совокупности. При таком предположении распределение случайных величин (2) совпадает с распределением случайных величин $\mu(0), \mu(1), \dots, \mu(i)$, равных числам частиц, существовавших в процессе Гальтона–Ватсона $\mu(k), k = 0, 1, \dots$, определяемого соотношениями

$$\mu(0) = 1, \quad \mathbf{E} u^{\mu(1)} = \mathbf{E} u^\eta.$$

Производящую функцию $\mathbf{E} u^\eta$ обозначим символом $\varphi(u)$.

Заметим теперь, что величина $|\overline{D}(h, i)|$ является последним элементом в цепочке

$$|\overline{D}(h, t)|, |\overline{D}(h, t-1)|, \dots, |\overline{D}(h, i)|, \quad (4)$$

причем каждый следующий элемент цепочки выражается через предыдущий элемент цепочки с помощью соотношений

$$|\overline{D}(h, j)| = \sum_{y \in \overline{D}(h, j+1)} |\overline{D}(h, y, j+1)|, \quad j = t-1, \dots, i, \quad (5)$$

где

$$\overline{D}(h, y, j) = \{z \in D(y, j, j) \mid H(\lambda_{j,t}(z)) \leq h\}.$$

Пусть

$$z \in D(y, j, j), \quad y \in \overline{D}(h, j+1) \quad \text{и} \quad k = t - (j-1).$$

Тогда верны соотношения

$$H(\lambda_{j,t}(z)) = H(z, \lambda_{j+1,t}(y)) = \varphi_k(z) \parallel H(\lambda_{j+1,t}(y)),$$

$$H(\lambda_{j+1,t}(y)) \leq h.$$

Предположим, что каждый элемент $z \in D(y, j, j)$ независимо от других элементов с вероятностью $p_k = 2^{-n_k}$ удовлетворяет соотношению

$$\varphi_k(z) \parallel H(\lambda_{j+1,t}(y)) \leq h$$

и с вероятностью $q_k = 1 - p_k$ ему не удовлетворяет. Тогда распределение случайной величины $|\overline{D}(h, y, j)|$ совпадает с распределением случайной величины $\varsigma_k = \sum_{i=1}^n \xi_{k,i}$, где $\xi_{k,i}, i = 1, 2, \dots$, — последовательность независимых одинаково распределенных случайных величин, принимающих значения 1 и 0 с вероятностями p_k и q_k соответственно. Производящая функция $\Psi_k(u)$ случайной величины ς_k имеет вид

$$\Psi_k(u) = \varphi(q_k + p_k u).$$

Первая величина в цепочке (4) равна

$$|\overline{D}(h, t)| = |\{x \in X \mid \varphi_0(x) \leq h\}|.$$

Применяя описанный выше подход, предположим, что она имеет такое же распределение, как случайная величина $\varsigma_0 = \sum_{i=1}^N \xi_{0,i}$.

Пренебрежем зависимостью между случайными величинами вида $|\overline{D}(h, y, j)|$ в формуле (5): будем считать, что они независимы в совокупности. В этом случае распределение случайных величин (4) совпадает с распределением случайных величин $\bar{\mu}(0), \bar{\mu}(1), \dots, \bar{\mu}(t-i)$, равных числам частиц, существовавших в неоднородном ветвящемся процессе $\bar{\mu}(k), k = 0, 1, \dots$, у которого производящая функция числа потомков частицы в момент времени k равна $\Psi_k(u)$, а производящая функция $\Psi_0(u)$ числа начальных частиц равна $(q_0 + p_0 z)^N$.

Таким образом, в используемой теоретико-вероятностной модели будем предполагать, что при любых $x \in X$ и $i = 1, \dots, t$:

- случайные величины $|D(x, i)|$ и $|\overline{D}(h, i)|$ независимы,
- распределение случайной величины $|D(x, i)|$ при любом $x \in X$ совпадает с распределением случайной величины $\mu(i)$,
- распределение случайной величины $|\overline{D}(h, i)|$ при любом значении маркера $h \in V(i)$ совпадает с распределением случайной величины $\bar{\mu}(t-i)$.

Теперь, в соответствии с формулой (1) и введенной теоретико-вероятностной моделью, мы можем получить значение средней временной сложности построения дополнительных цепочек при полной обработке одной таблицы RMM-метода:

$$\begin{aligned}
 T_{RMM} &= \mathbf{E} T_2 = \sum_{i=1}^t \frac{mi}{N^2} \sum_{h \in V(i)} \mathbf{E} \sum_{x \in \overline{D}(h,i)} |\overline{D}(h,i)| \mathbf{E} |D(x,i)| = \\
 &= \sum_{i=1}^t \frac{mi}{N^2} \sum_{h \in V(i)} \mathbf{E} |\overline{D}(h,i)|^2 = \\
 &= \sum_{i=1}^t \frac{mi}{N} \frac{1}{N} \sum_{h \in V(i)} \mathbf{E} \bar{\mu}^2(t-i) = \\
 &= \sum_{i=0}^{t-1} \frac{m(t-i)}{N} \frac{1}{N} \prod_{j=0}^i p_j \mathbf{E} \bar{\mu}^2(i). \tag{6}
 \end{aligned}$$

Описанный выше подход с небольшими изменениями пригоден также для вычисления сложности построения дополнительных цепочек в RM-методе. В этом случае распределение процесса $\bar{\mu}(i), i = 0, 1, \dots$, совпадает с распределением процесса $\mu(i), i = 0, 1, \dots$, и средняя сложность построения дополнительных цепочек при полной обработке одной таблицы RM-метода описывается формулой

$$T_{RM} = \sum_{i=0}^{t-1} \frac{m(t-i)}{N} \mathbf{E} \mu^2(i). \tag{7}$$

4. Сложность построения дополнительных цепочек с использованием и без использования маркеров

Для вычисления величин T_{RMM} и T_{RM} необходимо знать средние значения $\mathbf{E} \bar{\mu}^2(i)$ и $\mathbf{E} \mu^2(i)$. Известно (см. [21]), что для критического процесса Гальтона – Ватсона, начинающегося с одной частицы и имеющего производящую функцию распределения числа непосредственных потомков одной частицы $\varphi(u)$, $\varphi''(1) < \infty$, справедлива формула

$$\mathbf{E} \mu^2(i) = 1 + \varphi''(1) i, \quad i = 0, 1, \dots, t-1. \tag{8}$$

При аналогичных ограничениях на производящую функцию $\varphi(u)$ значения величин $\mathbf{E} \bar{\mu}^2(i)$, $i = 0, 1, \dots, t-1$, описывает следующая теорема.

Теорема 1. Пусть $\bar{\mu}(i)$, $i = 0, 1, \dots$, — неоднородный ветвящийся процесс размножения и гибели частиц с дискретным временем и случайным числом начальных частиц. Если производящая функция числа потомков одной частицы в момент времени i имеет вид $\Psi_i(u) = \varphi(q_i + p_i u)$, $q_i = 1 - p_i$, а производящая функция числа начальных частиц имеет вид $(q_0 + p_0 u)^N$, $q_0 = 1 - p_0$, то

$$\mathbf{E} \bar{\mu}^2(i) = N \left(\prod_{j=0}^i p_j \right) \left((N-1) \prod_{j=0}^i p_j + 1 + \varphi''(1) \sum_{k=1}^i \prod_{j=k}^i p_j \right) \quad (9)$$

для всех $i = 0, 1, \dots, t-1$.

Доказательство. Рассмотрим последовательность производящих функций $f_i(u) = \mathbf{E} u^{\bar{\mu}(i)}$, $i = 0, 1, \dots$. Ее члены связаны рекуррентными соотношениями

$$f_0(u) = (q_0 + p_0 u)^N, \quad f_i(u) = f_{i-1}(\Psi_i(u)), \quad i = 1, 2, \dots \quad (10)$$

Дифференцируя соотношения (10) в точке $u = 1$, находим рекуррентные соотношения между первыми моментами:

$$\mathbf{E} \bar{\mu}(0) = N p_0, \quad \mathbf{E} \bar{\mu}(i) = \mathbf{E} \bar{\mu}(i-1) \cdot \Psi'_i(1), \quad i = 1, 2, \dots$$

Дифференцируя соотношения (10) дважды в точке $u = 1$, получаем рекуррентные соотношения между первыми и вторыми моментами:

$$\mathbf{E} \bar{\mu}^{[2]}(0) = N^{[2]} p_0^2,$$

$$\mathbf{E} \bar{\mu}^{[2]}(i) = (\mathbf{E} \bar{\mu}^{[2]}(i-1) + \mathbf{E} \bar{\mu}(i-1) \varphi''(1)) p_i^2, \quad i = 1, 2, \dots$$

Отсюда находим, что верны равенства

$$\mathbf{E} \bar{\mu}(i) = N \prod_{j=0}^i p_j,$$

$$\mathbf{E} \bar{\mu}^{[2]}(i) = N^{[2]} \prod_{j=0}^i p_j^2 + N \varphi''(1) \left(\prod_{j=0}^i p_j \right) \left(\sum_{k=1}^i \prod_{j=k}^i p_j \right),$$

$$\mathbf{E} \bar{\mu}^2(i) = N \left(\prod_{j=0}^i p_j \right) \left((N-1) \prod_{j=0}^i p_j + 1 + \varphi''(1) \left(\sum_{k=1}^i \prod_{j=k}^i p_j \right) \right).$$

Доказательство завершено. $\square$

Подстановка формул (8), (9) в формулы (6), (7) позволяет получить точные выражения для величин T_{RMM} и T_{RM} в рассматриваемой модели. Однако точные формулы слишком сложны для практических вычислений, поэтому приведем оценки этих величин.

Теорема 2. Пусть

$$v(x) = \prod_{0 \leq j \leq tx} p_j, \quad \alpha = \frac{mt}{N},$$

а величины T_{RMM} и T_{RM} задаются формулами (6) и (7) соответственно. Тогда верны равенства

$$\frac{T_{RMM}}{\alpha t^2} = \varphi''(1) \int_0^1 \int_0^x (1-x) \frac{v(x)}{v(y)} dy dx + \int_0^1 (1-x) v(x) \frac{N}{t} dx + \varepsilon_{RMM}, \quad (11)$$

$$\frac{T_{RM}}{\alpha t^2} = \frac{\varphi''(1)}{6} + \varepsilon_{RM}, \quad (12)$$

где для величин ε_{RMM} и ε_{RM} справедливы оценки

$$0 \leq \varepsilon_{RMM} \leq \frac{1}{t} \left(\frac{N}{t} \int_0^1 v(x) dx + 1 + \varphi''(1) \right), \quad -\frac{\varphi''(1)}{6t^2} \leq \varepsilon_{RM} \leq \frac{1}{t}.$$

Доказательство. Вычислим величину $T_{RMM}/(\alpha t^2)$. Для этого подставим формулу (8) в формулу (6) и выполним цепочку преобразований:

$$\begin{aligned} \frac{T_{RMM}}{\alpha t^2} &= \frac{N}{mt^3} \sum_{i=0}^{t-1} \frac{m(t-i)}{N} \left((N-1) \prod_{j=0}^i p_j + 1 + \varphi''(1) \sum_{k=1}^i \prod_{j=k}^i p_j \right) = \\ &= \sum_{i=0}^{t-1} \left(1 - \frac{i}{t} \right) \left(\frac{1}{t} \left((N-1) \prod_{j=0}^i p_j + 1 \right) + \varphi''(1) \sum_{k=1}^i \frac{1}{t} \prod_{j=k}^i p_j \right) \frac{1}{t} = \\ &= \int_0^1 \left(1 - \frac{[xt]}{t} \right) \left(\frac{1}{t} \left((N-1) v(x) + 1 \right) + \varphi''(1) \int_0^x \frac{v(x)}{v(y)} dy \right) dx. \end{aligned}$$

Отсюда получаем представление

$$\begin{aligned} \frac{T_{RMM}}{\alpha t^2} &= \int_0^1 (1-x) \left(v(x) \frac{N}{t} + \varphi''(1) \int_0^x \frac{v(x)}{v(y)} dy \right) dx + \varepsilon_{RMM}, \\ \varepsilon_{RMM} &= \int_0^1 \left(x - \frac{[xt]}{t} \right) v(x) \frac{N}{t} dx + \int_0^1 \left(1 - \frac{[xt]}{t} \right) \frac{1}{t} (1-v(x)) dx + \\ &\quad + \varphi''(1) \int_0^1 \left(x - \frac{[xt]}{t} \right) \int_0^x \frac{v(x)}{v(y)} dy dx. \end{aligned}$$

Учитывая, что при любых x, y из области интегрирования выполняются оценки

$$0 \leq x - \frac{[xt]}{t} \leq \frac{1}{t}, \quad 0 \leq 1 - v(x) \leq 1, \quad \frac{v(x)}{v(y)} \leq 1,$$

получаем

$$0 \leq \varepsilon_{RMM} \leq \frac{1}{t} \left(\frac{N}{t} \int_0^1 v(x) dx + 1 + \varphi''(1) \right).$$

Вычислим величину $T_{RMM}/(\alpha t^2)$. Для этого подставим формулу (9) в формулу (7) и выполним цепочку преобразований:

$$\begin{aligned} \frac{T_{RMM}}{\alpha t^2} &= \sum_{i=0}^{t-1} \left(\left(1 - \frac{i}{t} \right) \cdot \left(\frac{1}{t} + \varphi''(1) \cdot \frac{i}{t} \right) \right) \frac{1}{t} = \\ &= \frac{1}{2t} \left(1 + \frac{1}{t} \right) + \frac{\varphi''(1)}{6} \left(1 - \frac{1}{t^2} \right) = \frac{\varphi''(1)}{6} + \varepsilon_{RM}, \\ \varepsilon_{RM} &= \frac{1}{2t} \left(1 + \frac{1}{t} \left(1 - \frac{\varphi''(1)}{3} \right) \right). \end{aligned}$$

Отсюда находим, что

$$-\frac{\varphi''(1)}{6t^2} \leq \varepsilon_{RM} \leq \frac{1}{2t} \left(1 + \frac{1}{t} \right) \leq \frac{1}{t}. \quad \square$$

5. Один подход к решению задачи оптимального выбора проверочных точек

Из результатов теоремы следует, что при больших значениях параметра t выигрыш T_{RMM}/T_{RM} в сложности построения дополнительных цепочек определяется суммой двух слагаемых

$$\alpha_1(v) = \frac{\varphi''(1)}{\varphi''(1)/6} \int_0^1 \int_0^x (1-x) \frac{v(x)}{v(y)} dy dx = 6 \int_0^1 \int_0^x (1-x) \frac{v(x)}{v(y)} dy dx$$

и

$$\alpha_2(v) = \frac{1}{\varphi''(1)/6} \int_0^1 (1-x) \left(v(x) \frac{N}{t} \right) dx.$$

Изучим задачу минимизации этих слагаемых.

Пусть множество позиций проверочных точек I_0 состоит из величин $i_0, \dots, i_{s-1}$, причем $0 = i_0 < i_1 < \dots < i_{s-1} < t$, $\sum_{j=1}^{s-1} n_{i_j} = r$.

Свяжем с этим набором неубывающую последовательность $x_1, x_2, \dots, x_r$ из r точек интервала $(0, 1)$. Первые n_{i_1} точек положим равными i_1/t , следующие n_{i_2} точек положим равными i_2/t и так далее.

Эта последовательность точек определяет разбиение отрезка $[0, 1]$ на $r + 1$ последовательных интервалов $\Lambda_0 = [0, x_1)$, $\Lambda_j = [x_j, x_{j+1})$, $j = 1, \dots, r-1$, $\Lambda_r = [x_r, 1]$. Длины этих интервалов обозначим символами Δ_j , $j = 0, \dots, r$ (некоторые интервалы могут иметь нулевую длину).

С учетом введенных обозначений получаем

$$v(x) = \prod_{0 \leq j \leq tx} p_j = \sum_{i=0}^r \frac{I\{x \in \Lambda_i\}}{2^{n_0+i}}, \quad \frac{v(x)}{v(y)} = \sum_{i=0}^r \sum_{j=0}^i \frac{I\{x \in \Lambda_i, y \in \Lambda_j\}}{2^{i-j}}.$$

С помощью этих соотношений вычисляем слагаемое

$$\begin{aligned} \alpha_1(v) &= 6 \int_0^1 \int_0^x \frac{v(x)}{v(y)} (1-x) dy dx = \\ &= 6 \sum_{i=0}^r \sum_{j=0}^i \frac{1}{2^{i-j}} \int_0^1 \int_0^x I\{x \in \Lambda_i, y \in \Lambda_j\} (1-x) dy dx = \\ &= 6 \sum_{i=0}^r \int_0^1 \int_0^x I\{x \in \Lambda_i, y \in \Lambda_i\} (1-x) dy dx + \\ &\quad + 6 \sum_{i=0}^r \sum_{j=0}^{i-1} \frac{1}{2^{i-j}} \int_0^1 \int_0^x I\{x \in \Lambda_i, y \in \Lambda_j\} (1-x) dy dx = Q_r(\bar{\Delta}), \end{aligned}$$

где

$$Q_r(\bar{\Delta}) = 3 \sum_{i=0}^r \Delta_i^2 \left(\sum_{j=i+1}^r \Delta_j + \frac{1}{3} \Delta_i \right) + 6 \sum_{i=1}^r \sum_{j=0}^{i-1} \frac{\Delta_i \Delta_j}{2^{i-j}} \left(\sum_{k=i+1}^r \Delta_k + \frac{1}{2} \Delta_i \right).$$

Следовательно, задача выбора позиций проверочных точек, для которых слагаемое $\alpha_1(v)$ близко к минимальному значению, связана с задачей нахождения точки минимума функции $Q_r(\bar{y})$, определенной на гиперплоскости $\sum_{i=0}^r y_i = 1$. В частности, подходящее значение вектора $\bar{\Delta}$ следует выбирать как можно ближе к точке минимума функции $Q_r(\bar{y})$ на множестве G , ограниченном условиями $y_0 > 0$, $y_r > 0$, $y_i \geq 0$, $i = 1, \dots, r-1$.

Минимум функции $Q_r(\bar{y})$ на G обозначим символом a_r . Он достигается либо в точке локального минимума, либо на границе множества G .

Поиск локального минимума функции $Q_r(\bar{y})$ с помощью метода множителей Лагранжа (см. [19]) сводится к решению системы уравнений второй степени нелинейности от $r + 2$ переменных. Общие методы решения систем второй степени нелинейности в настоящее время не известны. Тем не менее найти точку локального минимума функции $Q_r(\bar{y})$ с любой наперед заданной точностью можно с помощью универсальных вычислительных методов поиска экстремума функции нескольких переменных.

Применимость этих методов обеспечивается тем, что функция $Q_r(\bar{y})$ имеет в G не более одной точки локального минимума (значение функции $Q_r(\bar{y})$ на любой прямой из гиперплоскости $\sum_{i=0}^r y_i = 1$ описывается полиномиальной функцией степени не выше 3 и, следовательно, не может иметь на ней более одного локального минимума).

Поиск минимума функции $Q_r(\bar{y})$ на границах множества G сводится к изучению локальных минимумов этой функции на подмножествах множества G , получаемых обнулением части переменных $y_1, \dots, y_{r-1}$. Вычисление локальных минимумов на всех 2^{r-1} , $r = 1, \dots, 21$, вариантах подобных подмножеств методом координатного спуска [20], показало, что все они превышают локальный минимум из внутренней точки множества G . Следовательно, при всех $r = 1, \dots, 21$ минимум функции $Q_r(\bar{y})$ достигается в точке локального минимума.

В табл. 1 на с. 113 приводятся полученные методом координатного спуска приблизительные значения величин a_r и точек локального минимума $\bar{y} = (y_0, \dots, y_r)$, $r = 1, 2, \dots, 21$.

Таким образом, если $\sum_{i=1}^{s-1} n_{j_i} = r$, то в качестве набора позиций проверочных точек, на котором значение $\alpha_1(v)$ близко к минимальному, можно предложить набор из r позиций однобитных проверочных точек вида

$$i_j = \left[t \left(\sum_{k=0}^{j-1} y_k \right) \right], \quad j = 1, \dots, r, \quad (13)$$

где величины $y_0, \dots, y_{r-1}$ приведены в r -й строке табл. 1 (при $r = 1, \dots, 21$).

При изучении влияния проверочных точек на слагаемое $\alpha_2(v)$ воспользуемся его представлением в виде трех сомножителей

$$\alpha_2(v) = \left(p_0 \frac{N}{t} \right) \frac{1}{\varphi''(1)} 6C(v), \quad C(v) = \int_0^1 \frac{v(x)}{p_0} (1-x) dx.$$

Первый множитель

$$p_0 \frac{N}{t} = \frac{N}{2^{n_0} t}$$

за счет увеличения n_0 может быть сделан достаточно малым. Вторым множителем $\frac{1}{\varphi''(1)}$ для большинства отображений G можно полагать равным 1. Это объясняется тем, что для случайного равновероятного отображения G величина $\varphi''(1)$ с ростом N по вероятности стремится к 1. Третий множитель $6C(v)$ при любом выборе позиций проверочных точек не превосходит 3. Его приближенные значения b_r , $r = 1, \dots, 21$, для наборов проверочных точек из табл. 1 приведены в табл. 2.

Таблица 1. Приблизительные значения величин a_r и точек минимума $y_0, \dots, y_r$

r	a_r	$y_0, y_1, \dots, y_r$
0	1.0000	1.000
1	0.7778	0.333, 0.667
2	0.6401	0.258, 0.162, 0.581
3	0.5454	0.212, 0.124, 0.145, 0.519
4	0.4759	0.180, 0.102, 0.113, 0.131, 0.473
5	0.4225	0.157, 0.087, 0.094, 0.104, 0.121, 0.435
6	0.3802	0.140, 0.076, 0.081, 0.088, 0.097, 0.113, 0.405
7	0.3457	0.126, 0.068, 0.072, 0.076, 0.082, 0.091, 0.106, 0.379
8	0.3171	0.114, 0.061, 0.064, 0.068, 0.072, 0.078, 0.086, 0.100, 0.358
9	0.2930	0.105, 0.056, 0.058, 0.061, 0.064, 0.068, 0.073, 0.081, 0.094, 0.339
10	0.2723	0.097, 0.051, 0.053, 0.055, 0.058, 0.061, 0.065, 0.070, 0.077, 0.090, 0.323
11	0.2544	0.090, 0.048, 0.049, 0.051, 0.053, 0.055, 0.058, 0.062, 0.067, 0.074, 0.086, 0.308
12	0.2387	0.084, 0.044, 0.046, 0.047, 0.049, 0.051, 0.053, 0.056, 0.059, 0.064, 0.071, 0.082, 0.295
13	0.2249	0.079, 0.041, 0.043, 0.044, 0.045, 0.047, 0.049, 0.051, 0.054, 0.057, 0.061, 0.068, 0.079, 0.283
14	0.2126	0.075, 0.039, 0.040, 0.041, 0.042, 0.043, 0.045, 0.047, 0.049, 0.052, 0.055, 0.059, 0.065, 0.076, 0.273
15	0.2016	0.071, 0.037, 0.038, 0.038, 0.039, 0.041, 0.042, 0.043, 0.045, 0.047, 0.050, 0.053, 0.057, 0.063, 0.073, 0.263
16	0.1917	0.067, 0.035, 0.035, 0.036, 0.037, 0.038, 0.039, 0.041, 0.042, 0.044, 0.046, 0.048, 0.051, 0.055, 0.061, 0.071, 0.254
17	0.1827	0.064, 0.033, 0.034, 0.034, 0.035, 0.036, 0.037, 0.038, 0.039, 0.041, 0.042, 0.044, 0.047, 0.049, 0.053, 0.059, 0.069, 0.246
18	0.1746	0.061, 0.031, 0.032, 0.033, 0.033, 0.034, 0.035, 0.036, 0.037, 0.038, 0.039, 0.041, 0.043, 0.045, 0.048, 0.052, 0.057, 0.066, 0.239
19	0.1671	0.058, 0.030, 0.030, 0.031, 0.032, 0.032, 0.033, 0.034, 0.035, 0.036, 0.037, 0.038, 0.040, 0.042, 0.044, 0.047, 0.050, 0.056, 0.065, 0.232
20	0.1603	0.055, 0.029, 0.029, 0.030, 0.030, 0.031, 0.031, 0.032, 0.033, 0.034, 0.035, 0.036, 0.037, 0.039, 0.040, 0.043, 0.045, 0.049, 0.054, 0.063, 0.225
21	0.1540	0.053, 0.027, 0.028, 0.028, 0.029, 0.029, 0.030, 0.031, 0.031, 0.032, 0.033, 0.034, 0.035, 0.036, 0.038, 0.039, 0.041, 0.044, 0.048, 0.052, 0.061, 0.219

Таблица 2. Приближенные значения множителя $6C(v)$

r	0	1	2	3	4	5	6	7	8	9	10
b_r	3.000	2.333	1.920	1.636	1.428	1.267	1.141	1.037	0.951	0.879	0.817
r	11	12	13	14	15	16	17	18	19	20	21
b_r	0.763	0.717	0.675	0.638	0.605	0.575	0.548	0.524	0.501	0.481	0.462

6. Результаты экспериментальной проверки

С целью сравнения полученных в теореме оценок средней сложности T_{RMM} и выборочного среднего величины $T(a)$, $a \in X$, были проведены две серии экспериментов.

Каждая серия состояла из 9 однотипных экспериментов, которые отличались между собой значением параметра r , $r = 1, 2, \dots, 9$. В каждом эксперименте для однонаправленной функции G , заданной на множестве битовых векторов длины $n = 26$, строилась радужная таблица из $m = N^{2/3}$, $N = 2^n$, цепочек длины $t = N^{1/3}$. При построении таблицы использовался маркер цепочки с r однобитными проверочными точками, позиции которых определялись формулой (13) (оптимальное расположение проверочных точек). Редуцированная конечная точка состояла из $n_0 = n - r$ бит. В эксперименте случайным образом генерировалось $D = 250$ элементов данных a_i , $i = 1, \dots, D$, для которых подсчитывалось число F -операций $T(a_i)$, вычисленных при построении дополнительных цепочек на всех t шагах оперативного этапа RMM-метода, используемого для обращения G в точке a_i . По окончании эксперимента подсчитывалось выборочное среднее

$$S(r) = \frac{1}{D} \sum_{i=1}^D T(a_i).$$

Так как величина T_{RMM} является асимптотической оценкой среднего значения $T(a)$, то следует ожидать, что при больших значениях D отношение $S(r)/T_{RMM}$ близко к 1.

В соответствии с полученными выше результатами при больших значениях N, t величины

$$\frac{T_{RMM}}{\alpha t^2 \varphi''(1)/6} \quad \text{и} \quad a_r + \left(\frac{1}{2^{n-r}} \frac{N}{t} \right) \frac{1}{\varphi''(1)} b_r$$

приблизительно равны. Следовательно, можно ожидать, что приблизительно равны величины

$$\gamma_{\text{экс}}(r) = \frac{S(r)}{\alpha t^2 \varphi''(1)/6} \quad \text{и} \quad \gamma_{\text{теор}}(r) = a_r + \left(\frac{1}{2^{n-r}} \frac{N}{t} \right) \frac{1}{\varphi''(1)} b_r.$$

В следующих двух таблицах для $r = 1, \dots, 9$ приведены значения $\gamma_{\text{экс}}(r)$ и $\gamma_{\text{теор}}(r)$. В табл. 3 отражены результаты эксперимента с функцией G , у которой выполнено соотношение $\varphi''(1) \approx 1$. В табл. 4 отражены результаты аналогичного эксперимента с функцией G при $\varphi''(1) \approx 2$.

Таблица 3. Значения $\gamma_{\text{экс}}(r)$ и $\gamma_{\text{теор}}(r)$ при $\varphi''(1) \approx 1$

r	1	2	3	4	5	6	7	8	9
$\gamma_{\text{теор}}(r)$	0.789	0.659	0.577	0.533	0.522	0.560	0.673	0.915	1.401
$\gamma_{\text{экс}}(r)$	0.789	0.658	0.587	0.537	0.534	0.572	0.679	0.917	1.416

Таблица 4. Значения $\gamma_{\text{экс}}(r)$ и $\gamma_{\text{теор}}(r)$ при $\varphi''(1) \approx 2$

	1	2	3	4	5	6	7	8	9
$\gamma_{\text{теор}}(r)$	0.783	0.650	0.562	0.504	0.472	0.470	0.509	0.617	0.846
$\gamma_{\text{экс}}(r)$	0.778	0.650	0.563	0.506	0.480	0.478	0.517	0.615	0.862

Анализ таблиц показывает, что полученные в работе оценки при достаточно больших значениях параметров хорошо соответствуют экспериментальным данным, что говорит об адекватности использованных в работе вероятностных предположений.

Заключение

В заключение сделаем следующие выводы.

1. Множество позиций проверочных точек маркера, позволяющего достичь наибольший выигрыш в средней сложности построения дополнительных цепочек при полной обработке одной таблицы, состоит из редуцированной n_0 -битной конечной точки и r однобитных проверочных точек.
2. Величины n_0 и r подбираются из соображений минимизации суммы

$$a_r + \frac{b_r}{\varphi''(1)} \cdot \frac{N}{2^{n_0 t}},$$

где величины a_r указаны в табл. 1, а величины b_r в табл. 2 при $r = 0, \dots, 21$. Значение данной суммы приближает отношение $\frac{T_{RMM}}{T_{RM}}$ средних сложностей построения дополнительных цепочек при полной обработке одной таблицы в RMM- и RM-методах. Для однонаправленной функции G , случайно и равновероятно выбранной из множества всех функций $X \rightarrow X$, величина

$$\varphi''(1) = \sum_{k=1}^N k(k-1) \cdot \frac{|\{x \mid |G^{-1}(x)| = k\}|}{N}$$

с высокой вероятностью близка к 1.

3. Расположение однобитных проверочных точек описывается формулой (13) посредством величин $y_0, \dots, y_r$, указанных в табл. 1.

Список литературы

- [1] Hellman M., “A cryptanalytic time-memory trade-off”, *IEEE Trans. Inf. Theory*, **IT-26**:4 (1980), 401–406.
- [2] Denning D., *Cryptography and Data Security*, Boston, MA: Addison-Wesley, 1982, xiv+400 pp.
- [3] Oechslin P., “Making a faster cryptanalytic time-memory trade-off”, *Lect. Notes Comput. Sci.*, **2729**, 2003, 617–630.
- [4] Biryukov A., Shamir A., Wagner D., “Real time cryptanalysis of A5/1 on a PC”, *Lect. Notes Comput. Sci.*, **1978**, 2000, 1–18.
- [5] Saarinen M.-J. O., “A time-memory tradeoff attack against LILI-128”, *Lect. Notes Comput. Sci.*, **2365**, 2001, 231–236.
- [6] Mentens N., Batina L., Preneel B., Verbaauwhede I., “Time-memory trade-off attack on FPGA platforms: UNIX password cracking”, *Lect. Notes Comput. Sci.*, **3985**, 2006, 323–334.
- [7] Bono S., Green M., Stubblefield A., Juels A., Rubin A., Szydlo M., “Security analysis of a cryptographically-enabled RFID device”. In: “*14th USENIX Security Symposium. USENIX’05*”, Berkeley, CA: USENIX Assoc., 2005, 1–15.
- [8] Avoine G., Junod P., Oechslin P., “Characterization and improvement of time-memory tradeoff based on perfect tables”, *ACM Trans. Inf. Syst. Security (TISSEC)*, **11**:4 (2008), 1–22, 17.
- [9] Avoine G., Junod P., Oechslin P., “Time-memory trade-offs: False alarm detection using checkpoints”, *Lect. Notes Comput. Sci.*, **3797**, 2005, 183–196.
- [10] Barkan E., Biham E., Shamir A., “Rigorous bounds on cryptanalytic time/memory tradeoffs”, *Lect. Notes Comput. Sci.*, **4117**, 2006, 1–21.
- [11] Avoine G., Bourgeois A., Carpent X., “Analysis of rainbow tables with fingerprints”, *Lect. Notes Comput. Sci.*, **9144**, 2015, 356–374.
- [12] Hong J., “The cost of false alarms in Hellman and rainbow tradeoffs”, *Designs, Codes and Cryptography*, **57**:3 (2010), 293–327.
- [13] Wang W., Lin D., “Analysis of multiple checkpoints in non-perfect and perfect rainbow tradeoff revisited”, *Lect. Notes Comput. Sci.*, **8233**, 2013, 288–301.
- [14] Kim J.W., Seo J., Hong J., Park K., Kim S.-R., “High-speed parallel implementations of the rainbow method in a heterogeneous system”, *Lect. Notes Comput. Sci.*, **7668**, 2012, 303–316.
- [15] Kim J.W., Seo J., Hong J., Park K., Kim S.R., “High-speed parallel implementations of the rainbow method based on perfect tables in a heterogeneous system.”, *Software: Practice and Experience*, **45**:6 (2015), 837–855.
- [16] Lee G.W., Hong J., *A comparison of perfect table cryptanalytic tradeoff algorithms*, 2012. <http://eprint.iacr.org>.
- [17] Hong J., Moon S., “A comparison of cryptanalytic tradeoff algorithms”, *J. Cryptology*, **26**:4 (2013), 559–637.
- [18] Hong J., “Perfect rainbow tradeoff with checkpoints revisited”, *PLoS ONE*, **11**:11 (2016), 18, e0166404.
- [19] Зорич В. А., *Математический анализ. Часть 1. Изд. 2-е*, М.: ФАЗИС, 1997, 518 с.
- [20] Корн Г., Корн. Т., *Справочник по математике для научных работников и инженеров*, М.: Наука, 1970, 832 с.
- [21] Севастьянов Б. А., *Ветвящиеся процессы*, М.: Наука, 1971, 436 с.