УДК 519.72 (075.8)

МОДЕЛИ НА ОСНОВЕ РЮКЗАЧНОГО ВЕКТОРА С ОБРАТНЫМ ПРЕОБРАЗОВАНИЕМ

Осипян В. О.[1]***, Подколзин В. В.***[2]

MODELS WITH BASIS OF KNAPSACK VECTOR WITH INVERSE TRANSFORMATION

Osipyan V. O., Podkolzin V. V.

Models of systems on the basis of unilateral transformations are investigated. Models are built on the basis of knapsack vector problems with inverse transformation. Properties of the offered models are analyzed.

Keywords: a knapsack vector, variation, injectiveness, sequence of values, unilateral function, model, transformation system

Необходимость контроля аппаратно-канальных изменений в системах передачи, хранения и обработки информации, а также темпы развития программно-аппаратных средств определяют непрерывный рост интереса к математическому моделированию задач теории и практики передачи, обработки и защиты информации. В основе большинства задач перечисленных областей лежит односторонняя функция. Под односторонней функцией понимается отображение, для которого поиск обратного отображения либо связан с какой-то NP-полной задачей, либо эффективный алгоритм его нахождения еще не известен.

Задача о рюкзаке для заданных $w \in N$ и вектора $A = (a_1, a_2, \ldots, a_n)$, где $a_i \in N$, $i = 1, \ldots, n$, имеет решение в Z_p, если существует решение x уравнения

$$Ax^T = w, x \in Z_p^n, \qquad (1)$$

где $Z_p = \{0, \ldots, p-1\}$ — кольцо вычетов по модулю p, $Z_p^n = \{x \mid x = (x_1, \ldots, x_n), x_i \in Z_p, i = 1, \ldots, n\}$.

Решение уравнение (1) можно рассматривать как некоторое отображение $F_{\text{пр}} : W \to X$, где $W = \{w \,|\, w \in N\}$, $X = \{x \,|\, x \in Z_p^n\}$. Функцию $F_{\text{пр}}$ будем называть функцией, представляющей прямое решение задачи о рюкзаке. С другой стороны, если в (1) в качестве исходных данных рассматривать A и x, то имеем отображение $F_{\text{об}} : X \to W$, представляющее обратное решение задачи о рюкзаке. $F_{\text{об}}$ — умножение двух векторов. Вектор A уравнения (1) будем называть рюкзачным вектором.

Обозначим через $\mu_p(A)$ множество значений w, для которых уравнение (1) имеет решение в Z_p, т.е. множество всех различных допустимых числовых значений, которые могут быть представлены в виде линейной комбинации компонентов вектора A с коэффициентами из Z_p.

В основе многих методов поиска данных с использованием хеш-функций, хеш-таблиц, декартова дерева, фильтра Блума, в системах защиты информации лежат отображения числовых значений во множество числовых цепочек (или чисел с заданными свойствами). Так как $F_{\text{пр}}$ и $F_{\text{об}}$ представляют собой аналогичные функции, можно слелать вывод о целесообразности моделирования отображений на основе задачи о рюкзаке, область применения которой лежит в теории кодирования, защиты информации, алгоритмизации, WEB-программировании, баз данных. Сложность задачи нахождения функции отображения в моделях на основе задачи о рюкзаке зависит, прежде всего, от свойств рюкзачного вектора.

[1]Осипян Валерий Осипович, д-р физ.-мат. наук, профессор кафедры информационных технологий Кубанского государственного университета; e-mail: rrwo@mail.ru

[2]Подколзин Вадим Владиславович, старший преподаватель кафедры информационных технологий Кубанского государственного университета; e-mail: vvp_35@mail.ru

В последнее время широкое распространение получили системы с преобразованием исходных значений в двоичные последовательности. Т. е. для рюкзачных систем заданный двоичный вектор x длины n отображается в значение w, которое получено скалярным умножением x на рюкзачный вектор A согласно (1). Очевидно, что в этом случае $w \in \mu_p(A)$, следовательно уравнение (1) имеет решение. Если A — инъективный вектор, то решение единственно. При такой модели функционирования системы отображения для заданного рюкзачного вектора операция определения w является примитивной, а обратный процесс в случае инъективного, но не сверхрастущего рюкзачного вектора может потребовать дополнительных ресурсов.

Определим модель отображения, в которой процесс восстановления исходных значений будет простым и способен соответствовать нахождению скалярного произведения, а основной объем вычислений будет выполняться на этапе нахождения значений соответствующих исходным данным.

Рассмотрим модель с инъективным рюкзачным вектором [1] $A = (a_1, a_2, \ldots, a_n)$ в Z_p. Вариацией вектора A в Z_p назовем вектор $\Delta A = (\delta_1, \delta_2, \ldots, \delta_n)$, где $\delta_1 = a_1, \delta_i = a_i - \sum_{j=1}^{i-1}(p-1)a_j, i = 2, \ldots, n$. Обозначим через $max(A)$ значение наибольшего компонента вектора A.

Переведем исходную последовательность значений $(v_1, v_2, \ldots, v_l)$ в последовательность u-значных $(u_1, u_2, \ldots, u_l)$ чисел в k-ричной системе счисления $(u = [\log_k(\max(v_i))] + 1)$. Выпишем полученные значения в виде строки последовательности цифр u_i $(i = 1, \ldots, l)$ и разобьем полученную строку на части (блоки) длины m (в случае необходимости строка дополняется нулями), где m удовлетворяет соотношению

$$\sum_{i=1}^{m}(k-1)k^{i-1} \leqslant \sum_{j=1}^{n}(p-1)a_j < \sum_{i=1}^{m+1}(k-1)k^{i-1}.$$

Каждому блоку $(\beta_1, \beta_2, \ldots, \beta_m)$ поставим в соответствие числовое значение

$$v = \sum_{i=1}^{m}\beta_i k^{i-1},$$

которое принадлежит отрезку

$$\left[0, \sum_{i=1}^{m}(p-1)k^{i-1}\right].$$

Тогда существуют две точки w' и w'', $w' < w''$ такие, что $w' \in \mu_p(A)$, $w'' \in \mu_p(A)$ и

$$\forall\ w \in \mu_p(A)\ w \neq w', w \neq w'' \Rightarrow \Rightarrow w \notin [w', w''] \quad (2)$$

Т. е. найдутся два соседних значения w' и w'', что $v \in [w', w'']$. Следовательно

$$v = w' + \delta. \quad (3)$$

Здесь $\delta \leqslant \delta_{\max} = \max(\Delta A)$ [2]. В общем случае, в силу инъективности вектора A, имеем $\delta \notin \mu_p(A)$ [3].

На основании вышеизложенного в качестве результата отображения для произвольного значения исходного текста v определим вектор

$$x_v = (\alpha_1, \alpha_2, \ldots, \alpha_n, \eta_0, \eta_1, \ldots, \eta_s),$$

в том числе

$$\sum_{i=1}^{n}\alpha_i a_i = w', \quad \sum_{j=0}^{s}\eta_j p^j = \delta,$$

где $s = [\log_p(\max(\Delta A))] + 1$.

Заметим, что первые n компонентов вектора x_v соответствуют коэффициентам разложения w по рюкзачному вектору, а остальные $s+1$ — представлению δ в p-ричной системе счисления. При таком определении отображения можно сделать вывод о том, что условие (2) не является обязательным.

Действительно, для отрезка $[w_i, w_{i+1}]$ (w_i, w_{i+1} — два последовательных значения $W_{\mu_p(A)}$ [3]), что $v \in [w_i, w_{i+1}]$ выполняется $v = w_i + \delta$, где $\delta \leqslant \delta_{\max} = \max(\Delta A)$.

В случае, если рюкзачный вектор A является инъективным, но не сверхрастущим, ΔA — его вектор вариации, и существует $j \leqslant n-2$, так что два значения последовательности значений w_{j+2} и w_j образуют отрезок, содержащий v, то представление (3) может быть не единственным. Но в любом случае размер вектора значений x_v не превышает $n + s + 1$.

В силу того, что для любого рюкзачного вектора A плотность $d_p(A) \leqslant 1$ [3], значение s может быть большим. Например, для рюкзачного вектора $A = (5, 16, 41, 98, 228, 528, 1225, 2855, 6697, 15820, 37651, 90296, 218250, 531736, 1306016, 3234182)$ в Z_2 для двоичного представления исходного текста длина блока $m = 23$. Так как, $\Delta A = (5, 11, 20, 36, 68, 140, 309, 714, 1695, 4133, 10144,$

25138, 62796, 158032, 400576, 1022627), то $s = [\log_2(1022627)] + 1 = 20$. Таким образом, длина соответствующего значения одного блока в Z_2 равна 36.

Заметим, что данная модель преобразования, для которой $s = 0$ (т. е. если отбросить часть не являющуюся коэффициентами разложения по компонентам вектора), тоже допустима. В частности, данная модель полностью соответствует требованиям, предъявляемым к отображениям хеширования.

Модифицируем метод получения значения отображения с целью уменьшения его длины.

Определим разложение произвольного значения ω по компонентам вектора $A = (a_1, a_2, \ldots, a_n)$ следующим образом:

$$\alpha_n = \omega \operatorname{div} a_n,$$
$$\alpha_i = (\omega - \sum_{j=i+1}^{n} \alpha_j a_j) \operatorname{div} a_i, \alpha_0 = \omega - \sum_{j=1}^{n} \alpha_j a_j. \quad (4)$$

Из (4) следует $\alpha_i \leqslant (a_{i+1} - 1) \operatorname{div} a_i = \rho_i$, $i = 1, \ldots, n-1$ (div — операция целочисленного деления).

Обозначим $\rho_n = \max(2, \rho_1, \ldots, \rho_{n-1})$. Определим основание Z_p следующим образом $p = \rho_n + 1$. Тогда коэффициенты $\alpha_1, \alpha_2, \ldots, \alpha_n$ разложения (4) произвольного числа из отрезка $[0, \sum_{j=1}^{n} \rho_j a_j]$ удовлетворяют соотношениям $\alpha_i \leqslant \rho_i$, $i = 1, \ldots, n$. Длина блока m в k-ричной кодировке определяется согласно

$$\sum_{i=1}^{m} (k-1)k^{i-1} \leqslant \sum_{j=1}^{n} \rho_j a_j < \sum_{i=1}^{m+1} (k-1)k^{i-1}.$$

В качестве значения отображения для произвольного блока исходного значений v определим вектор

$$x_v = (\alpha_1, \alpha_2, \ldots, \alpha_n, \eta_0, \eta_1, \ldots, \eta_s),$$

где α_i, $i = 1, \ldots, n$, определяются согласно (4), $s = [\log p(a_1 - 1)] + 1$ и выполняется

$$\sum_{j=0}^{s} \eta_j p^j = v - \sum_{j=1}^{n} \alpha_j a_j.$$

При таком способе определения отображения длина вектора x_v зависит, прежде всего, от значения первого компонента рюкзачного вектора. Рассмотрим только возрастающие рюкзачные вектора, так как компоненты, в которых нарушается строгая монотонность в разложении (4), будут иметь нулевые коэффициенты.

Применяя метод нахождения p для вектора $A = (5, 16, 41, 98, 228, 528, 1225, 2855, 6697, 15820, 37651, 90296, 218250, 531736, 1306016, 3234182)$ из предыдущего примера, находим что $p = 4$. Т. е. получаем, что длина блока $m = 23$, а длина вектора значений равна 18 в Z_4. Если обратиться к способам представления вектора $x_v = (x_1, x_2, \ldots, x_{18})$ в памяти ЭВМ, то можно выделить два основных способа: поэлементную кодировку или кодировку числового значения. В первом случае передаются 18 двоичных чисел длины 2, соответствующие двоичному представлению каждого компонента x_v, для чего требуется 36 битов памяти. Во втором случае передается двоичное значение числа $\sum_{i=1}^{18} x_i 4^{i-1} \leqslant 3 \sum_{j=1}^{18} 4^{j-1}$, для хранения которого необходимо 33 бита. Возможны комбинации этих двух способов. Например, передавать двоичные представления двух чисел $\sum_{i=1}^{9} x_i 4^{i-1}$ и $\sum_{i=10}^{18} x_i 4^{i-10}$. В этом случае потребуется $18 + 18 = 36$ битов памяти, т. е. результат отображения можно передавать несколькими значениями меньшего размера.

Заметим, что поэлементная кодировка тоже может быть модифицирована с целью оптимизации размера памяти, используемой для хранения. Например, для рюкзачного вектора $A = (5, 16, 41, 98, 228, 528, 1225, 2855, 6697, 15820, 37651, 90296, 218250, 531736, 1306016, 53234182)$ имеем $\rho_1 = 3$, $\rho_2 = \cdots = \rho_{n-2} = 2$, $\rho_{n-1} = 40$. Поэтому для хранения x_{n-1}, x_n можно выделить 6 битов памяти, а для остальных x_i ($i = 1 \ldots n-2$) по 2 бита. Таким образом, А является универсальным рюкзаком [4] и $s = [\log_{\max(2,\rho_i)}(a_1 - 1)] + 1$. Преимуществом последнего метода преобразования является отсутствие информации о рюкзачном векторе во множестве значений.

Несмотря на большие затраты памяти по сравнению с ранее приведенным примером в Z_2, следует отметить ряд преимуществ данной модели отображения:

1) значение p не постулируется в системе, а определяется на основе рюкзачного вектора;

2) количество компонентов s результирующего значения, не являющихся коэффициентами разложения по компонетнам рюкзачного вектора, оно зависит только от его первого компонента;

3) при поэлементной кодировке длина представления различных компонентов исходной последовательности значений может быть различна;

4) разделение значения отображения на части при смешанном кодировании.

При решении задач восстановления рюкзачного вектора, полученных методами, основанными на обратной задаче о рюкзаке, допустимы только статистические методы или метод на основе выбранного текста. Использование последних неэффективно для систем предложенной модели. В моделях на основе прямой задачи о рюкзаке значение отображения w заведомо принадлежит $\mu_p(A)$, следовательно, предоставляет информацию о неких свойствах рюкзачного вектора. При использовании систем на основе обратной задачи аналогичная информация может быть получена только в случае, если известна пара значений (исходное и результирующее) и количество s элементов вне разложения по компонентам рюкзачного вектора. Преимуществом данной модели является тот факт, что для получения значений и однозначного восстановления исходной информации не требуется выполнения условия инъективности рюкзачного вектора, а лишь условия возрастания значений его компонентов. Данный факт является существенным, особенно для моделей с динамически генерируемыми векторами, так как алгоритм проверки инъективности рюкзачного вектора требует времени $O(\Pi_{\delta_i<0}|\delta_i|)$ (не для сверхрастущего случая). Затраты при прямом и обратном преобразовании для данного метода составляют $O(n)$.

Модификация модели позволяет определить вероятностную ее модификацию. Пусть $A = (a_1, a_2, \ldots, a_n)$ — произвольный рюкзачный вектор в Z_p, v — числовое значение блока исходной последовательности. Тогда соответствующее значение $x_v = (\alpha_1, \alpha_2, \ldots, \alpha_n, \eta_0, \eta_1, \ldots, \eta_s)$ можно определить следующим образом:

1) $i = n$;

2) определить значение α_i как произвольное число из отрезка $[0, \min(p-1, v \operatorname{div} \mathrm{a}_i)]$;

3) $v = v - \alpha_i \mathrm{a}_i$;

4) $i = i - 1$;

5) если $i > 0$, то перейти к п.2);

6) найти значения разрядов $\eta_0, \eta_1, \ldots, \eta_s$ числа v в p-ричной системе счисления;

7) x_v определить как $(\alpha_1, \alpha_2, \ldots, \alpha_n, \eta_0, \eta_1, \ldots, \eta_s)$.

Основным преимуществом данного метода преобразования, наряду с произвольностью вектора A, является «непредсказуемость» получаемого результата. Данное обстоятельство не влияет на процесс восстановления исходных значений, но для одного и того же исходного текста позволяет получить различные результаты отображения.

Вероятностная модель имеет два существенных недостатка. Во-первых, большой размер векторов значений, так как $s = [\log_p n] + 1 = \text{const}$ (в случае, когда все i определятся как 0). Во-вторых, при большом значении размера применимости или возможности задавать входные данные, в условии $s = \text{const}$, нелегальный пользователь может определить достаточное количество значений из $\mu_p(A)$ (правда, неизвестно за какое время), что позволит ему восстановить рюкзачный вектор.

Произвольную всюду определенную функцию $F : R^k \to Z_p^n$, где R^k — пространство векторов длины n вещественных компонентов, назовем генератором векторов размерности n ($\Gamma\mathrm{B}_n$) в Z_p. Вектор $\mathrm{A} = (\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_n)$ определяется генератором векторов F, если существует $\lambda \in R^k$, что $F(\lambda) = \mathrm{A}$. В качестве $\Gamma\mathrm{B}_n$ могут выступать алгоритм, аналитическая функция или их совокупность, важным здесь является то, что $F(\lambda)$ может быть найдено за приемлемое (в том или ином смысле) время.

Если вектор $F(\lambda) = \mathrm{A}$, то A можно рассматривать как рюкзачный вектор размерности n. С другой стороны, возможна интерпретация $F(\lambda)$ как вариация вектора $\Delta\mathrm{A}$, на основе которого возможно получение вектора A. В обоих случаях считаем, что $F(\lambda)$ задает вектор A и обозначим $F(\lambda) \Rightarrow \mathrm{A}$.

Определим модель с динамически определяемым рюкзачным вектором на основе обратной задачи о рюкзаке:

1) $\lambda' \in R^q$ является частью системы, недоступной для сторонних пользователей;

2) генератор векторов $F : R^k \to Z_p^n$ является общедоступным внешним исходным данным;

3) информация о способе кодирования последовательности исходных значений является общедоступной;

4) значение $t \in N$ является общедоступным и определяется в качестве параметра отображения;

5) $j = 1$;

6) значение $\lambda_j \in R^l (q + l = k)$ является общедоступным и определяется в качестве параметра отображения;

7) определяется рюкзачный вектор $F(\lambda', \lambda_j) \Rightarrow$ A, на основании которого определяются p, m — длина входных данных и длина соответствующего значения $n + s + 1$;

8) для очередного блока v_i можно найти соответствующее значение w_i согласно (4);

9) $j = j + 1$;

10) если $j \leqslant t$, то перейти к п. 6;

11) повторять п.п. 5–10 последовательно для всех блоков входных данных;

12) последовательность $(t, \lambda_1, w_1, \ldots, w_t, \lambda_2, w_{t+1}, \ldots, w_{2t}, \ldots, w_s)$ является результатом отображения последовательности входных числовых значений.

Рассмотрим задачи восстановления рюкзачного вектора при данной модели отображения. Анализ множества значений заведомо неэффективен, так как для генератора векторов с длиной числовых значений l необходимо рассмотреть для каждого блока C_{n+l-2}^{n+1} разбиений. Для каждого разбиения определить основание системы счисления и получить коэффициенты разложения. На основе полученных коэффициентов и значения блока исходных данных (если таковой известен) перебрать возрастающие вектора длины n. Используя каждый вектор как возможный рюкзачный вектор, предпринять попытку найти параметры генератора векторов. Каждая из задач, которые приходиться решать, является NP-полной задачей. Например, для векторов размерности n с длиной значений $l = 66$ в среднем нужно просмотреть порядка $5 \cdot 10^{16}$ вариантов разбиений. Пусть выбрано разбиение, при котором $s = 5$ и имеются 8 частей по 2 бита, 4 части — по 3 бита, 1 — 5 битов, 2 — 6 битов, 1 — 16 битов. Тогда неоходимо просмотреть более $2{,}6 \cdot 10^{14}$ различных комбинаций оснований систем счисления для данного разбиения. Объем описанных вычислений говорит о технической невыполнимости задачи поиска рюкзачного вектора. Вследствие вышеописанных причин единственный вариант нахождения функции отображения данной модели — поиск значения $\lambda' \in R^q$. Результат такого поиска существенно зависит от свойств функции генератора векторов и в общем случае является малоэффективным даже для достаточно простых функций.

Литература

1. *Саломаа А.* Криптография с открытым ключом. М.: Мир, 1995. 318 с.
2. *Подколзин В. В., Осипян В. О.* Верхняя граница числа решений обобщенной задачи о рюкзаке на заданном входе // Актуальные проблемы безопасности информационных технологий: Материалы III Междунар. научно-практич. конф. С. 28-32. Красноярск: Сибирский гос. аэрокосмич. ун-т, 2009. 142 с.
3. *Подколзин В. В., Осипян В. О.* О свойствах рюкзачных систем защиты информации с открытым ключом в Zp // Вестник сибирского гос. аэрокосмич. ун-та им. акад. М. Ф. Решетнева. С. 51–55. Красноярск, 2010. Вып. 3 (29). 244 с.
4. *Осипян В. О.* Асимметрическая система защиты информации на основе универсального и функционального рюкзаков // Защита информации. Конфидент. 2004. №6. С. 61–63.

Ключевые слова: рюкзачный вектор, вариация, инъективность, последовательность значений, односторонняя функция, модель, система преобразования

Статья поступила 10 декабря 2010 г.
Кубанский государственный университет, г. Краснодар