УДК 519.72 (075.8)

ХЕШИРОВАНИЕ НА ОСНОВЕ ФУНКЦИОНАЛЬНОГО ГЕНЕРАТОРА РЮКЗАЧНЫХ ВЕКТОРОВ

Подколзин В. В., Лейман А. В., Панкова А. В.

HASHING ON THE BASIS OF THE FUNCTION GENERATOR OF KNAPSACK VECTORS

Podkolzin V. V., Layman A. V., Pankova A. V.

Kuban State University, Krasnodar, 350040, Russia
e-mail: antonina.leyman@gmail.com

Abstract. The article deals with the definition of the hash function based on the knapsack problem. It is proposed to utilize the functionally defined knapsack generator vectors. The hashing algorithm uses a forward and backward pass of input data. On each pass the data are divided into blocks, each of which defines its own knapsack vector. The result of applying knapsack vector defines the value affecting not only the current block of hash values but also the calculation of the next block. The applicability of the model for the problems of hashing in various fields was analyzed.

Keywords: hash, knapsack vector, function dynamically generated knapsack vector, cryptographic resistance

Введение

Задача о стандартном рюкзаке для заданных $w \in N$ и вектора $\mathbf{A} = (a_1, a_2, \ldots, a_n)$, где $a_i \in N$, $i = 1, \ldots, n$ имеет решение, если существует двоичный вектор $\mathbf{x}$ длины n, удовлетворяющий соотношению

$$\mathbf{A}\mathbf{x}^{\mathrm{T}} = w. \qquad (1)$$

Решение уравнения (1) можно рассматривать как некоторое отображение $F_{\text{пр}} : X \to W$, где $W = \{w \,|\, w \in N\}$, X — подмножество множества двоичных векторов длины n. С другой стороны, если в (1) в качестве исходных данных рассматривать $\mathbf{A}$ и $\mathbf{x}$, то имеем отображение $F_{\text{об}} : W \to X$, представляющее решение обратной задачи о рюкзаке. Используя уравнение (1), можно моделировать отображения, область применения которых лежит в теории кодирования, алгоритмизации, WEB-программировании, теории баз данных и защиты информации. Впервые применение NP-полной задачи о рюкзаке в области криптографии с открытым ключом было предложено Р. Меркле и М. Хеллманом [1], дальнейшее развитие данное направление получило в работах В. Чора и Р. Райвеста [2].

Обозначим пространство двоичных и вещественных векторов длины n через 2^n и R^n соответственно.

Определение. Произвольную всюду определенную векторную функцию $\mathbf{F} : R^k \to 2^n$ будем называть функциональным генератором векторов размерности n (ГРВ^n).

Использование генератора рюкзачных векторов позволило объединить сложность NP-задач дискретной факторизации и задачи о рюкзаке, а также воспользоваться свойствами полиалфавитных систем для создания моделей преобразования информации. Использование рюкзачных векторов большого размера в явном виде определило широкое использование задачи о рюкзаке только в целях защиты информации. Введение ГРВ^n позволило перенести центр внимания с самого рюкзачного вектора на метод его определения, а простота задания и использова-

Подколзин Вадим Владиславович, канд. физ.-мат. наук, доцент кафедры информационных технологий Кубанского государственного университета; e-mail: vvp_35@mail.ru

Лейман Антонина Васильевна, аспирант кафедры информационных технологий Кубанского государственного университета; e-mail: antonina.leyman@gmail.com

Панкова Александра Викторовна, магистрант кафедры информационных технологий Кубанского государственного университета; e-mail: jenndy92@gmail.com

ния генератора векторов — расширить сферу применения рюкзачных преобразований [3].

В предлагаемой статье рассматриваются вопросы применения модели преобразования информации с генератором рюкзачного вектора на основе математических функций в целях определения хеш-функций. В качестве исходных данных рассматриваются двоичные числовые последовательности, которые также интерпретируются и как вектора, и как двоичные представления соответствующих чисел.

Хеширование — преобразование обрабатываемых данных с помощью некоторого метода с последующим использованием полученного образа вместо исходных данных. Особое место в теории занимают криптографически стойкие хеш-функций. Для того чтобы хеш-функция $H(\alpha)$ считалась криптографически стойкой, она должна удовлетворять следующим основным требованиям:

1) Хеш-функция $H(\alpha)$ должна применяться к блоку данных любой длины.
2) Хеш-функция $H(\alpha)$ создает выход фиксированной длины.
3) $H(\alpha)$ вычисляется за время, ограниченное некоторым полиномом для любого значения α.
4) Для любого данного значения хеш-кода $\mathbf{h}$ вычислительно невозможно найти α такое, что $H(\alpha) = \mathbf{h}$.
5) Вычислительно невозможно найти произвольную пару (α, β) такую, что $H(\alpha) = H(\beta)$.

Следует отметить, что существование необратимых хеш-функций, для которых теоретически невозможно вычисление какого-либо прообраза заданного значения хеш-функции, не доказано. Обычно нахождение обратного значения является лишь вычислительно сложной задачей, в частности, если и в ее основе лежит NP-проблема.

1. Хеширование на основе ГРВ^n

Пусть n — длина рюкзачного вектора. Длина хеш-значения m определяется соотношением $m = b * n, b \geqslant 1$. Здесь и далее полагается, что длина набора данных превышает m, в противном случае необходимо дополнить последовательность исходной информации до необходимой длины путем добавления последовательности $\tilde{\alpha}(|\tilde{\alpha}| \geqslant m - |\alpha|)$. Последнее может быть осуществлено либо повтором набора данных, либо добавлением к нему фиксированной последовательности. В зависимости от способа применения модели в ГРВ^n $\mathbf{F} : S \times P \to 2^n$ определение функции $\mathbf{F}$ может относится как к открытому, так и к закрытому ключу. Закрытый ключ S может отсутствовать.

Разобьём исходную последовательность данных $\alpha = \alpha_1\alpha_2 \dots \alpha_L$ ($\alpha_i \in \{0,1\}$, $i = 1, \dots, L$) длины $|\alpha| = L$ на t блоков по m элементов

$$\alpha = \alpha_1\alpha_2 \dots \alpha_m\alpha_{m+1}\alpha_{m+2} \dots \alpha_{2m} \dots \\ \dots \alpha_{(t-1)m+1}\alpha_{(t-1)m+2} \dots \alpha_{tm}\alpha_{tm+1}\alpha_{tm+2} \dots \\ \dots \alpha_{tm+k}.$$

В общем случае имеет место $L = tm + k$, $0 < k < m$. Обозначим i-й блок $\bar{\alpha}_i = \alpha_{im+1}\alpha_{im+2} \dots \alpha_{(i+1)m}$, $i = 1, \dots, t$. В свою очередь, каждый блок может быть разбит на подблоки длины n

$$\bar{\alpha}_i = \beta_1\beta_2 \dots \beta_n\beta_{n+1}\beta_{n+2} \dots \beta_{2n} \dots \beta_{bn} \\ (\beta_i \in \{0,1\}, \quad i = 1, \dots, L).$$

Обозначим через $\bar{\beta}_i^j$ j-й подблок блока $\bar{\alpha}_i$, $j = 1 \dots, b$. Для каждого блока определяется собственный рюкзачный вектор, последовательно применяемый для подблоков как прямое преобразование $F_{\text{пр}}$ по модулю 2^n. Старшие биты, не используемые в преобразовании, суммарно определяют значение рюкзачного вектора следующего блока. Для выведения результата прямой задачи о рюкзаке и области значений рюкзачного преобразования предлагается использовать поразрядную инверсию в зависимости от числа Хэмминга. Для обеспечения влияния последних k бит на результирующее хеш-значение необходимо выполнить двунаправленную обработку входных данных.

2. Алгоритм вычисления хеш-значения на основе ГРВ^n

Хеш-значение для входной последовательности α определяется как результат применения последовательности динамически генерируемых рюкзачных векторов на основе заданной функции $\mathbf{F}$ для каждого блока $\bar{\beta}_i^j$ и последующего сложения полученных значений по модулю 2, как представлено в описанном далее алгоритме.

Алгоритм 1: ХешФГРВ

INPUT: α, s, r, m, $\mathbf{F}: S \times P \to 2^{\mathbf{m}}$
OUTPUT: $w = w_1 w_2 \dots w_b$ — m-битное хеш-значение

1. если $|\alpha| <= m$, то $\alpha \leftarrow \alpha\tilde{\alpha}$
2. для $u = 1 \dots b$ выполнить $w_u \leftarrow 0$
3. $p \leftarrow 0$
4. $t \leftarrow |\alpha| : m$
5. для $i = 1 \dots t$ выполнить ($\bar{\alpha}_i = \beta_1\beta_2 \dots$ $\dots \beta_m = \bar{\beta}_i^1 \bar{\beta}_i^2 \dots \bar{\beta}_i^b$)
   5.1. $p \leftarrow (p + \sum_{d=1}^{r} 2^{d-1}\beta_d) \bmod 2^r$
   5.2. $\mathbf{A} \leftarrow \mathbf{F}(s, p)$
   5.3. для $j = 1 \dots b$ выполнить
      5.3.1. $\gamma \leftarrow \mathbf{A}\bar{\beta}_i^{j\mathrm{T}}$
      5.3.2. если $H(\gamma) \bmod 2 = 1$ выполнить $\gamma \leftarrow \neg\gamma$
      5.3.3. $w_j \leftarrow w_j \oplus (\gamma \bmod 2^n)$
      5.3.4. $p \leftarrow p + \gamma : 2^n$
6. $\alpha \leftarrow \overleftrightarrow{\alpha}$
7. повторить п. 5

В алгоритме ХешФГРВ используются следующие обозначения: $\tilde{\alpha}$ — двоичная последовательность длины более $m - |\alpha|$, такая, что $|\alpha - \tilde{\alpha}| \neq 0 (\bmod m)$; $\leftarrow$ — операция присваивания; : – операция целочисленного деления; mod — остаток от деления целых чисел; $\oplus$ — операция побитного сложения по модулю 2; $\neg$ — операция побитного отрицания; $\overleftrightarrow{\alpha} = \alpha_L \alpha_{L-1} \dots \alpha_1$; $H(x)$ — функция Хэмминга.

Опишем алгоритм 1 хеширования на основе функционально определенного генератор рюкзачного вектора размерности n. Входными параметрами алгоритма ХешФГРВ являются α — входная двоичная последовательность длины L; $\mathbf{F}: S \times P \to 2^n$ — функция, определяющая ГРВ^n; s — закрытый ключ ГРВ^n; r — длина двоичного вектора открытого ключа p ГРВ^n; m — длина двоичного хеш-значения, $m \equiv 0 (\bmod n)$.

Первый шаг обеспечивает длину входной последовательности не менее длины результирующего значения путем дописывания, в случае необходимости, двоичной цепочки $\tilde{\alpha}$. Значение $\tilde{\alpha}$ определяется конкретной реализацией данного алгоритма и может представлять собой несколько раз повторенную либо фиксированную двоичную последовательность, а также входное α.

Второй шаг инициализирует хеш-значение w нулем, w представляется как последовательность $b = m : n$ двоичных цепочек $w_1 w_2 \dots w_b$, $|w_i| = n$, $i = 1, \dots, b$.

Третий шаг определяет начальное значение P открытого ключа ГРВ^n.

На четвертом шаге вычисляется количество t последовательных блоков $\bar{\alpha}_i$ длины m цепочки α.

Пятый шаг повторяется для каждого блока $\bar{\alpha}_i$ и состоит в последовательном выполнении шагов 5.1–5.3.

Шаг 5.1 изменяет текущее значение открытого ключа P на числовое значение, представленное первыми r битами обработанного на четвертом шаге блока $\bar{\alpha}_i$ по модулю 2^r.

На шаге 5.2 на основе заданной функции $F(s, p)$ определяется рюкзачный вектор $\mathbf{A}$ [4].

Шаг 5.3 описывает очередную итерацию изменения хеш-значения w. Для каждого подблока $\bar{\beta}_i^j$ текущего блока $\bar{\alpha}_i$ последовательно выполняются пп. 5.3.1–5.3.4 алгоритма.

На шаге 5.3.1 вычисляется значение γ путем применения рюкзачного вектора $\mathbf{A}$ к текущему подблоку $\bar{\beta}_i^j$.

На шаге 5.3.2 в случае нечетного значения функции Хэмминга $H(\gamma)$ проводится поразрядное отрицание γ.

Шаг 5.3.3 определяет модификацию подблока w_j искомого хеш-значения w согласно последним n битам γ.

На шаге 5.3.4 значение открытого ключа P изменяется на число, представленное старшими $[\lg(\gamma)] - n$ битами значения γ.

Шестой шаг алгоритма инвертирует исходную последовательность и на седьмом шаге повторяется п. 5.

Алгоритм ХешФГРВ при выполнении пятого шага осуществляет последовательный поиск хеш-значения поблочно, что влечет исключение из рассмотрения последних k бит исходной последовательности. Последний факт противоречит свойствам лавинного эффекта и криптографической стойкости хеш-функций. Инвертирование исходной последовательности и повторение пятого шага позволяет устранить данный недостаток.

Понятие открытого ключа p ГРВ^n в контексте предложенного алгоритма является весьма условной и представляет собой истинно «открытое» значение только для первого блока исходной последовательности. Определение функционального генератора векторов с использованием секретного ключа не позволяет говорить о слабости алгоритма даже

на первом шаге. Для каждого последующего блока значение P изменяется в силу того, что максимальное величина области значений $\mu(\mathbf{A})$ рюкзачного вектора превышает 2^n [5].

3. Свойства хеширования на основе ГРВ^n

Задача восстановления исходной последовательности выходит за рамки NP-задачи о рюкзаке в силу того, что в общем случае $w \notin \mu(\mathbf{A})$ [6], кроме того, используется более одного рюкзачного вектора (шаг 5.2 Алгоритма 1). Наличие секретного ключа s определяется возможностью использования предлагаемой модели построения хеш-функции в целях криптографии и в общем случае не является необходимой.

Возникновение коллизий определяется только на этапе имплементации функции ГРВ^n и зависит от ее свойств, то есть выходит за рамки анализа свойств предлагаемой модели.

Оценим возможность восстановления модели преобразования для заданной последовательности α и хеш-значения $H(\alpha)$. Рассмотрим самый простой случай, когда функция $\mathbf{F}$ известна и обратима, сложность определяется лишь параметром S, $|\alpha| = m + 1$, $m = n$. Тогда Алгоритм 1 может быть описан следующим образом.

Алгоритм2: ОдноблочныйХешФГРВ

INPUT: α, s, r, m, $\mathbf{F}: \ S \times P \to 2^m$
OUTPUT: w — m-битное хеш-значение

1. $w \leftarrow 0$
2. $(\bar{\alpha} = \alpha_1\alpha_2 \dots \alpha_m = \beta_1\beta_2 \dots \beta_m = \bar{\beta})$
3. $p \leftarrow (\sum_{d=1}^{r} 2^{d-1}\beta_d) \bmod 2^r$
4. $\mathbf{A} \leftarrow \mathbf{F}(s, p)$
5. $\gamma \leftarrow \mathbf{A}\bar{\beta}^T$
6. если $H(\gamma) \bmod 2 = 1$ выполнить $\gamma \leftarrow \neg\gamma$
7. $w \leftarrow w \oplus (\gamma \bmod 2^m)$
8. $p \leftarrow p + \gamma : 2^m$
9. $\alpha \leftarrow \alpha_{m+1}\alpha_m \dots \alpha_1$
10. $(\bar{\alpha} = \alpha_{m+1}\alpha_m \dots \alpha_2 = \beta_1\beta_2 \dots \beta_m = \bar{\beta})$
11. $p \leftarrow (p + \sum_{d=1}^{r} 2^{d-1}\beta) \bmod 2^r$
12. $\mathbf{A} \leftarrow F(s, p)$
13. $\gamma \leftarrow \mathbf{A}\bar{\beta}^T$
14. если $H(\gamma) \bmod 2 = 1$ выполнить $\gamma \leftarrow \neg\gamma$
15. $w \leftarrow w \oplus (\gamma \bmod 2^m)$

Следует отметить, что преобразование первого блока определяется только начальными r битами и значениями параметра s, которые задают первый рюкзачный вектор $\mathbf{A}$.

Пусть γ_5 и γ_{13} — значения γ, получаемые в п. 5 и п. 13, а $\delta_1\delta_2 \dots \delta_v\delta_{v+1} \dots \delta_{v+m}$ и $\phi_1\phi_2 \dots \phi_u\phi_{u+1} \dots \phi_{u+m}$ — их соответствующие двоичные представления, $w = (\gamma_5 \bmod 2^m) \oplus (\gamma_{13} \bmod 2^m)$. Следовательно, прежде чем искать значение s, необходимо определить $u + v + 2m$ бит γ_5 и γ_{13}. С учетом возможных инверсий (на шагах 6 и 14) имеем $2^{u+v+2m+4}$ вариантов, а количество решений для каждого варианта — $C^m_{2^{v+m}}$. Из всех решений уравнения шага 5 необходимо выбрать те, которые определяют равенство шага 13 для заданной последовательности α. Таким образом, общее количество вариантов существенно превышает 2^{2m}. Так как в настоящее время практически не применяются хеш-значения длиной меньше 64, можно утверждать, что анализ предложенной модели на основе функционального ГРВ^n алгоритмически неразрешим.

Заключение

Предложенная в данной статье модель хеширования может быть применена в схемах цифровой подписи, а также в качестве средства обеспечения аутентификации имитозащиты в протоколах аутентификации сообщений. Анализ (поиск методов нахождения обратного преобразования) данной модели требует одновременного решения нескольких независимых вычислительно трудных задач. Следовательно, вероятность появления эффективных способов ее компрометации существенно уменьшается. Использование последовательности рюкзачных векторов соответствует свойствам полиалфавитных отображений, что уменьшает эффективность ее частотного и статистического анализа. Использование функционального генератора рюкзачных векторов существенно упрощает использование NP-полной задачи о рюкзаке, так как не требует хранения векторов большой размерности.

Литература

1. *Merkle R., Hellman M.* Hiding information and signatures in trapdoor knapsacks // IEEE Transactions on Information Theory IT-24, 1978. P. 525–530.

2. *Chor B., Rivest R.* A knapsack-type public key cryptosystem based on arithmetic in finite fields // Advances in Cryptology, Crypto'84. Heidelberg etc.: Springer, 1985, p. 54-65; revised version in IEEE Trans. Inform. Theory IT-34, 1988. P. 901–909.
3. *Осипян В. О., Подколзин В. В.* Модели на основе рюкзачного вектора с обратным преобразованием // Экологический вестник научных центров Черноморского экономического сотрудничества. 2010. № 4. С. 59–63.
4. *Подколзин В. В., Осипян В. О.* Алгоритм построения инъективного возрастающего рюкзачного вектора // Математические методы и информационно-технические средства: труды V Всерос. науч.-практ. конф. Краснодар: Краснодарский ун-т МВД России, 2009. С. 141–145.
5. *Подколзин В. В., Осипян В. О.* Об одном методе определения верхней границы числа входов для рюкзачных систем защиты информации // Вестник Воронежского института МВД России. 2010. № 4. С. 83–90.
6. *Подколзин В. В.* Построение инъективных рюкзачных векторов на основе структурных и частотных свойств числовых множеств // Экологический вестник научных центров Черноморского экономического сотрудничества. 2010. № 4. С. 64–67.

References

1. Merkle R., Hellman M. Hiding information and signatures in trapdoor knapsacks. In *IEEE Transactions on Information Theory IT-24*, 1978, pp. 525–530.
2. Chor B., Rivest R. A knapsack-type public key cryptosystem based on arithmetic in finite fields. In: *Advances in Cryptology, Crypto'84*. Heidelberg etc.: Springer, 1985, p. 54–65; revised version in IEEE Trans. Inform. Theory IT-34, 1988, pp. 901–909.
3. Osipyan V. O., Podkolzin V. V. Modeli na osnove rjukzatchnogo vektora s obratnym preobrazovaniem [Models with basis of knapsack vector with inverse transformation]. *Ekologitchtsky vestnik nautchnyh cetrov Tchernomorskogo ekonomithceskogo sotrudnitchestva* [Ecological bulletin of research centers of the Black Sea Economic Cooperation], 2010, no. 4, pp. 59–63. (In Russian)
4. Podkolzin V. V., Osipyan V. O. Algoritm postroeniya iniektivnogo vozrastajushego rjukzatchnogo vektora [An algorithm for constructing an injective increasing knapsack vector]. *Trudy V Vseros. nautshn.-pract. konf. 'Matematitcheskie metody b informatciono-tekhnitcheskie sredstva', Krasnodar: Krasnodarsky un-t MVD Rossii* [Proc. of V All-Russia scientific and practical conf. 'Mathematical methods and information technology equipment', Krasnodar, Krasnodar univ. MOI Russia], 2009, pp. 141–145. (In Russian)
5. Podkolzin V. V., Osipyan V. O. Ob odnom metode opredeleniya verhnei granitcy tchisla vhodov dlya rjukzatchnyh system zachity informatcii [On a method of determining the upper limit of the number of inputs to knapsack security systems]. *Vestnik Voronezhskogo institute MVD Rossii* [Bulletin of Voronezh Institute of the Russian Interior Ministry], 2010, no 4, pp. 83–90. (In Russian)
6. Podkolzin V. V. Postroenie injektivnyh rjukzatchnyh vektorov na osnove strukturnyh i tchastotnyh svoistv tschislovyh mnozhestv [Construction of the injective knapsack vectors on the basis of the numerical sets structure and statistics properties]. *Ekologitchtsky vestnik nautchnyh cetrov Tchernomorskogo ekonomithceskogo sotrudnitchestva* [Ecological bulletin of research centers of the Black Sea Economic Cooperation], 2010, no 4, pp. 64–67. (In Russian)

Статья поступила 29 октября 2014 г.

© Подколзин В. В., Лейман А. В., Панкова А. В., 2014