Секция 7

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ В ДИСКРЕТНОЙ МАТЕМАТИКЕ

УДК 519.7 DOI 10.17223/2226308X/11/43

НОВЫЙ АЛГОРИТМ ПОРОЖДЕНИЯ ОСЛАБЛЯЮЩИХ ОГРАНИЧЕНИЙ В ЗАДАЧЕ ОБРАЩЕНИЯ ХЕШ-ФУНКЦИИ MD4-39[1]

И. А. Грибанова

Представлены результаты по обращению 39-шаговой версии криптографической хеш-функции MD4 (MD4-39). В их основе лежит специальный алгоритм генерации дополнительных ограничений, накладываемых на переменные сцепления, которые ослабляют исходную задачу поиска прообраза известного хеш-значения. Алгоритм осуществляет поиск ослабляющих ограничений через решение задачи оптимизации на булевом гиперкубе оценочной функции специального вида. При помощи разработанного алгоритма удалось найти новые ограничения на переменные сцепления, использование которых позволило построить атаку на MD4-39, время выполнения которой в десятки раз меньше, чем у лучшей известной атаки. С использованием найденных ограничений удаётся находить MD4-39-прообразы для примерно 65 % (в среднем) случайных 128-битных булевых векторов.

Ключевые слова: *криптографические хеш-функции, обращение хеш-функций, MD4, MD4-39, SAT.*

Хеш-функция MD4 [1] является одной из первых криптографических хеш-функций, построенных на базе конструкции Меркля — Дамгарда. Несмотря на то, что она была скомпрометирована по отношению к задаче поиска коллизий [2], MD4 остаётся стойкой к задаче обращения, в рамках которой требуется найти неизвестный 512-битный блок открытого текста, хеширование которого даёт известное хеш-значение. В данном контексте интерес представляют задачи обращения неполнораундовых версий хеш-функции MD4. Далее через MD4-k будем обозначать варианты MD4, использующие первые k ($k \leqslant 48$) шагов базового алгоритма.

До настоящего момента лучшей из реализованных на практике атак на MD4-k оставалась атака, описанная в [3]. В рамках этой атаки используются так называемые «условия Доббертина», предложенные в [4]. Основная идея состоит в наложении дополнительных ограничений на переменные сцепления на определённых шагах алгоритма вычиления хеша с целью вывода из этих ограничений некоторой информации, приводящей к быстрому решению уравнений криптоанализа. В [3] получаемая таким образом система уравнений криптоанализа сводится к задаче о булевой выполнимости (SAT) и решается при помощи SAT-решателя minisat. Основным достижением работы [3] является построение практических атак на MD4-k, $k \leqslant 39$. Следует, однако, отметить, что данные атаки оказались успешными лишь для некоторых конкретных хеш-значений, при этом решение одной задачи требует нескольких часов работы SAT-

[1]Работа поддержана Российским научным фондом, проект № 16-11-10046.

решателя. В [5] описан параллельный вариант атаки из [3], который, однако, не привёл к кардинальному улучшению достигнутых ранее результатов.

В настоящей работе представлены новые ослабляющие ограничения, которые позволяют существенно улучшить результаты из [3, 5]. Процесс поиска эффективных ослабляющих ограничений сведён к задаче оптимизации специальной оценочной функции на булевом гиперкубе. Во всех вычислительных экспериментах использовались пропозициональные кодировки, предусматривающие работу с «переменными переключения» [5]. Для построения этих кодировок применялась система Transalg [6, 7]. Оптимизация оценочной функции выполнялась метаэвристическим алгоритмом, относящимся к классу «Tabu search» (поиск с запретами) [8].

Кратко опишем основную идею метода автоматического поиска ослабляющих ограничений. Рассмотрим задачу обращения функции вида $f_{\text{MD4-}k} : \{0,1\}^{512} \to \{0,1\}^{128}$ при фиксированном k и сведём её к SAT. Пусть $C(f_{\text{MD4-}k})$ — шаблонная КНФ для данной задачи [7], X — множество всех булевых переменных в данной КНФ, а КНФ $C(f_{\text{MD4-}k}, \chi)$ является результатом подстановки в $C(f_{\text{MD4-}k})$ обращаемого хеш-значения $\chi \in \{0,1\}^{128}$.

Множество Q ослабляющих ограничений для рассматриваемой задачи — это множество $R = \{r_1, \ldots, r_Q\}$, в котором с каждым ограничением r_j, $j = 1, \ldots, Q$, связана булева переменная s_j из множества переменных переключения $S = \{s_1, \ldots, s_Q\}$, $S \cap X = \varnothing$. Произвольное ограничение r_j — это, как правило, формула следующего вида:

$$x_{j_1}^{\sigma_1} \wedge \ldots \wedge x_{j_{t_j}}^{\sigma_{t_j}}, \quad \{x_{j_1}, \ldots, x_{j_{t_j}}\} \subseteq X.$$

Рассмотрим следующую КНФ:

$$C_{r_j} = (\bar{s}_j \vee x_{j_1}^{\sigma_1}) \wedge \ldots \wedge (\bar{s}_j \vee x_{j_{t_j}}^{\sigma_{t_j}}).$$

Из формулы $s_j \wedge C_{r_j}$ по правилу единичной дизъюнкции (Unit Propagation rule, [9]) выводится формула $x_{j_1}^{\sigma_1} \wedge \ldots \wedge x_{j_{t_j}}^{\sigma_{t_j}}$. С другой стороны, применение поглощения к формуле $\bar{s}_j \wedge C_{r_j}$ даёт $\bar{s}_j$. В данной ситуации будем говорить, что ограничение r_j «активно» при $s_j = 1$ и «неактивно» при $s_j = 0$.

Область всевозможных наборов значений переменных из множества S — это $\{0,1\}^Q$. Таким образом, каждый ненулевой булев вектор $\lambda \in \{0,1\}^Q$ задаёт некоторый набор активных ослабляющих ограничений из множества R. Вопрос оценки эффективности конкретного набора ограничений из R нетривиален. На текущем этапе в качестве меры эффективности выбрана функция, заданная следующим образом.

В произвольном векторе $\lambda \in \{0,1\}^Q$ выделим множество компонент $\{\lambda_{h_1}, \ldots, \lambda_{h_d}\}$, $1 \leqslant d \leqslant Q$, равных единице. Данное множество задаёт набор активных ослабляющих ограничений с номерами $h_1, \ldots, h_d$. Рассмотрим функцию

$$\mu(\lambda) = \#\{x^\sigma : \tilde{C}(\lambda) \to_{\text{UP}} x^\sigma, x \in X^{\text{in}}\},$$

в которой $\tilde{C}(\lambda) = C(f_{\text{MD4-}k}, \chi) \wedge \bigwedge\limits_{j \in \{h_1, \ldots, h_d\}} C_{r_j}$, множество $X^{\text{in}} \subset X$ — это множество переменных в $C(f_{\text{MD4-}k}, \chi)$, кодирующих неизвестный 512-битный вход функции $f_{\text{MD4-}k}$, а запись «$\tilde{C}(\lambda) \to_{\text{UP}} x^\sigma$» означает, что из КНФ $\tilde{C}(\lambda)$ по правилу единичной дизъюнкции выведена переменная x либо её отрицание. Таким образом, значение функции $\mu(\lambda)$ равно числу формул вида x^σ над переменными из множества X^{in}, выводимых по правилу единичной дизъюнкции в результате активизации ослабляющих ограничений, заданных вектором λ.

Оценочная функция $\mu(\lambda)$ — это функция типа «черный ящик», аналитические свойства которой неизвестны. Для решения задачи поиска новых ослабляющих ограничений как задачи максимизации данной функции используется метаэвристический алгоритм, относящийся к классу алгоритмов поиска с запретами. В программной реализации алгоритма рассматриваются окрестности Хэмминга радиуса 1 в множестве $\{0,1\}^Q$.

С использованием описанного подхода получены новые результаты по обращению хеш-функции MD4-39. В частности, найдены два неизвестных ранее множества ослабляющих ограничений, заданные следующими наборами значений переменных переключения:

$$\rho_1: \quad 000000000000011011101110111010000000000$$
$$\rho_2: \quad 000000000000001011101110111011000000000$$

Эти ограничения позволяют не только эффективно обращать хеш-значения 0^{128} и 1^{128}, но и устойчиво находить MD4-39-прообразы для 128-битных векторов, сгенерированных случайным образом (таблица).

Поиск MD4-39 прообразов
для 500 случайно сгенерированных 128-битных векторов

Relaxation constraints	Сред. время решения, с	Макс. время решения, с	Кол-во задач (в % от общего числа)	
			Прообраз найден	Прообраз не найден
ρ_1	12	80	65	35
ρ_2	46	250	75	25

Для большинства задач (65–75 %) среднее время нахождения одного прообраза при помощи SAT-решателя minisat2.2 на одном ядре процессора Intel i7-3770K (3,5 GHz) составило меньше минуты. В то же время, используя условия из [3, 4], minisat2.2 не находит решения таких задач за несколько часов. Остальные задачи (25–35 %) соответствуют 128-битным векторам, для которых не существует MD4-39-прообразов, совместных с условиями ρ_1 или ρ_2, причём этот факт быстро доказывается SAT-решателем.

ЛИТЕРАТУРА

1. *Rivest R. L.* The MD4 message digest algorithm // LNCS. 1990. V. 537. P. 303–311.
2. *Wang X., Lai X., Feng D., et al.* Cryptanalysis of the hash functions MD4 and RIPEMD // LNCS. 2005. V. 3494. P. 1–18.
3. *De D., Kumarasubramanian A., and Venkatesan R.* Inversion attacks on secure hash functions using SAT solvers // LNCS. 2007. V. 4501. P. 377–382.
4. *Dobbertin H.* The first two rounds of md4 are not one-way // LNCS. 1998. V. 1372. P. 284–292.
5. *Gribanova I., Zaikin O., Otpuschennikov I., and Semenov A.* Using parallel SAT solving algorithms to study the inversion of MD4 hash function // Proc. Parallel Computational Technologies. 2017. P. 100–109.
6. *Отпущенников И. В., Семенов А. А.* Технология трансляции комбинаторных проблем в булевы уравнения // Прикладная дискретная математика. 2011. № 1. С. 96–115.
7. *Otpuschennikov I., Semenov A., Gribanova I., et al.* Encoding cryptographic functions to SAT using TRANSALG system // Proc. ECAI2016 — 22nd Europ. Conf. Artificial Intelligence. Hague, 2016. V. 285. P. 1594–1595.
8. *Glover F. and Laguna M.* TABU Search. Kluwer, 1999.
9. *Dowling W. F. and Gallier J. H.* Linear-time algorithms for testing the satisfiability of propositional horn formulae // J. Logic Programming. 1984. V. 1. No. 3. P. 267–284.