соответствий Галуа и свойств замкнутых множеств. Доказано также, что выводимости D_1, D_3, D_4, D_5 гарантируют сохранение поддержки: результатом применения их к строгим ассоциативным правилам с поддержкой не менее чем δ_0 всегда являются строгие ассоциативные правила с таким же порогом поддержки. Именно выводимости D_1, D_3, D_4, D_5 применяются в алгоритме MClose для распознавания избыточных строгих ассоциациативных правил и построения неизбыточного минимаксного базиса. Показано, что алгоритм MClose по времени работы сопоставим с алгоритмом Close. Между тем на практике он более чем в 2 раза уменьшает мощность минимаксного базиса, формируемого алгоритмом Close.

Подробное изложение представленных результатов можно найти в [8].

ЛИТЕРАТУРА

1. *Биркгоф Г., Барти Т.* Современная прикладная алгебра. СПб.: Лань, 2005. 400 с.
2. *Гуров С. И.* Булевы алгебры, упорядоченные множества, решетки: определения, свойства, примеры. М.: Книжный дом «ЛИБРОКОМ», 2013. 352 с.
3. *Батура Т. В.* Модели и методы анализа компьютерных социальных сетей // Программные продукты и системы. 2013. № 3. С. 130–137.
4. *Платонов В. В., Семенов П. О.* Методы сокращения размерности в системах обнаружения сетевых атак // Проблемы информационной безопасности. Компьютерные системы. 2012. № 3. С. 40–45.
5. *Кузнецов С. О.* Автоматическое обучение на основе анализа формальных понятий // Автоматика и телемеханика. 2001. № 10. С. 3–27.
6. *Zaki M. J and Hsiao C.-J.* Efficient algorithms for mining closed itemsets and their lattice structure // IEEE Trans. Knowledge Data Eng. 2005. V. 17. No. 4. P. 462–478.
7. *Майер Д.* Теория реляционных баз данных. М.: Мир, 1987. 608 с.
8. *Быкова В. В., Катаева А. В.* О неизбыточном представлении минимаксного базиса строгих ассоциативных правил // Прикладная дискретная математика. 2017. № 36. С. 113–126.

УДК 519.7 DOI 10.17223/2226308X/10/61

ОБРАЩЕНИЕ КРИПТОГРАФИЧЕСКИХ ХЕШ-ФУНКЦИЙ С ИСПОЛЬЗОВАНИЕМ НЕСБАЛАНСИРОВАННЫХ ПРИБЛИЖЕНИЙ РАУНДОВЫХ ФУНКЦИЙ[1]

И. А. Грибанова

Представлены результаты решения задач обращения неполнораундового варианта криптографической хеш-функции MD4 с использованием новой техники, которая включает в себя следующие этапы: замену некоторых раундовых подфункций MD4 несбалансированными булевыми функциями; решение полученной изменённой задачи; использование части информации из решения изменённой задачи для перехода к решению исходной задачи. Предлагаемая техника комбинируется с дополнительными условиями на переменные сцепления, введёнными ранее Г. Доббертином. Проведённые вычислительные эксперименты демонстрируют работоспособность предлагаемого подхода в применении к задаче обращения 39-шаговой версии MD4 (MD4-39).

Ключевые слова: *криптоанализ, обращение хеш-функций, MD4, SAT.*

[1]Работа поддержана грантом РНФ № 16-11-10046.

Хеш-функция MD4 [1] представляет собой один из первых примеров криптографических хеш-функций, построенных на основе конструкции Меркля — Дамгарда [2, 3]. В [4] данная функция была полностью скомпрометирована по отношению к атаке поиска коллизий. Однако для задачи обращения полнораундовой MD4 эффективных алгоритмов не предложено до сих пор. Насколько нам известно, все результаты по обращению неполнораундовых версий MD4 так или иначе основываются на идеях, высказанных Г. Доббертином [5]. В данной работе показано, что двухраундовый вариант MD4 (то есть вариант, включающий 32 шага базового алгоритма) не является стойким к атаке поиска прообраза известного хеша. Основная идея работы [5] состоит в использовании дополнительных условий, связывающих переменные сцепления на определённых шагах. Далее для условий данного типа мы применяем термин «условия Доббертина».

Насколько можно судить из открытых источников, наилучший на данный момент результат для рассматриваемой задачи — это нахождение за разумное время прообразов хешей, порождаемых 39-шаговой версией MD4, которую далее будем обозначать MD4-39. Впервые этот результат приведён в [6], где для решения задачи использован SAT-подход. Некоторый вариант условий Доббертина добавлялся в пропозициональную кодировку алгоритма в виде дополнительных ограничений. По утверждению авторов [6], на решение одной задачи уходит около 8 часов работы SAT-решателя minisat на ПК с весьма мощным на тот момент процессором. В работе [7] результаты по обращению 39-шагового варианта MD4 улучшены и описан процесс автоматического синтеза условий Доббертина при помощи параллельного алгоритма решения задачи о булевой выполнимости.

В настоящей работе представлены результаты по обращению неполнораундовых версий MD4 с использованием новой техники. Суть этой техники состоит в замене некоторых подфункций несбалансированными булевыми функциями, использование которых, по нашему мнению, должно делать задачу обращения соответствующего варианта рассматриваемой хеш-функции проще для SAT-решателя. Это предположение полностью подтверждается на практике. Будем называть проблему обращения варианта рассматриваемой хеш-функции, в котором подфункции некоторых раундовых функций заменены несбалансированными булевыми функциями, интерполированной задачей обращения.

Кратко опишем принцип построения несбалансированных булевых функций. Везде далее процесс перехода от исходной сбалансированной булевой функции к несбалансированной будем называть модификацией.

Напомним, что в MD4 вычисление значения переменной сцепления на i-м шаге описывается формулой

$$Q_i = (Q_{i-4} + \Phi_i(Q_{i-1}, Q_{i-2}, Q_{i-3}) + m_{p(i)} + k_i) <<< s_i, \quad i \in \{0, \ldots, 47\},$$

где Φ_i — раундовая функция i-го шага; $m_{p(i)}$ — 32-битный вектор, являющийся частью хешируемого сообщения; k_i, s_i — константы i-го шага. Заметим, что в алгоритме MD4 Q_i — это 32-битный булев вектор, и все операции выполняются над такими векторами побитово.

В экспериментах для построения пропозициональных кодировок использовалась система Transalg [8]. В данной системе трансляция в SAT процесса вычисления сложной булевой функции представляется в виде суперпозиции функций меньшей арности. К сожалению, в Transalg арность таких функций жестко задана, и на данном этапе удалось построить модификации только булевых функций арности 3.

Для задачи обращения MD4-39 осуществлялась модификация функций, задающих вектор Q_{34}. В табл. 1 приведены два примера таких функций (исходных и модифицированных), заданных таблицами (для различных компонент вектора Q_{34} использовались разные модификации).

Таблица 1

x	y	z	Исход.	Модиф.
0	0	0	0	1
0	0	1	1	1
0	1	0	0	0
0	1	1	1	1
1	0	0	0	0
1	0	1	0	0
1	1	0	1	1
1	1	1	1	1

x	y	z	Исход.	Модиф.
0	0	0	0	0
0	0	1	1	1
0	1	0	0	0
0	1	1	1	1
1	0	0	0	0
1	0	1	0	0
1	1	0	1	1
1	1	1	1	0

После решения интерполированной задачи требуется перейти к решению исходной задачи. С этой целью в исходную пропозициональную кодировку подставляется часть данных из решения интерполированной задачи. В роли таких данных используются части найденного хешируемого сообщения (всего 256 бит), на основе которых находится прообраз рассматриваемого хеш-значения.

В табл. 2 приведены результаты для задачи обращения 39-шаговой версии MD4, полученные с помощью программной реализации описанной выше техники. В первом

Таблица 2

Среднее время решения (по 30 тестам)

Решение исходной задачи без интерполяции	Переход к решению исходной задачи после интерполяции
2170 с	1670 с

столбце представлено среднее время решения рассматриваемой задачи в исходном варианте с использованием условий Доббертина [5, 6]. Отметим, что соответствующая интерполированная задача решалась в разы быстрее исходной, поэтому время её решения в расчёт не принималось. Во втором столбце представлено среднее время перехода от решения интерполированной задачи к решению исходной. Во всех вычислительных экспериментах использовался многопоточный SAT-решатель Plingeling [9], который запускался на одном рабочем узле кластера «Академик В. М. Матросов» [10] (32 ядра процессора Opteron 6276).

На данный момент полученные результаты выглядят весьма скромно. Основная причина этого в том, что используются несбалансированные функции очень маленькой арности — всего от трёх переменных. Для построения кодировок с несбалансированными функциями большей арности необходимо внести дополнительные изменения в систему Transalg. Мы надеемся это сделать в самое ближайшее время.

ЛИТЕРАТУРА

1. *Rivest R. L.* The MD4 message digest algorithm // LNCS. 1990. V. 537. P. 303–311.
2. *Merkle R. A.* Certified digital signature // LNCS. 1990. V. 435. P. 218–238.
3. *Damgard I. A.* A design principle for hash functions // LNCS. 1990. V. 435. P. 416–427.
4. *Wang X., Lai X., Feng D., et al.* Cryptanalysis of the hash functions MD4 and RIPEMD // LNCS. 2005. V. 3494. P. 1–18.

5. *Dobbertin H.* The first two rounds of md4 are not one-way // LNCS. 1998. V. 1372. P. 284–292.
6. *De D., Kumarasubramanian A., and Venkatesan R.* Inversion attacks on secure hash functions using SAT solvers // LNCS. 2007. V. 4501. P. 377–382.
7. *Gribanova I., Zaikin O., Otpuschennikov I., and Semenov A.* Using parallel SAT solving algorithms to study the inversion of MD4 hash function // Параллельные вычислительные технологии. XI Междунар. конф. ПаВТ'2017, г. Казань, 3–7 апреля 2017 г. Короткие статьи и описания плакатов. Челябинск: Издательский центр ЮУрГУ, 2017. С. 100–109.
8. *Otpuschennikov I., Semenov A., Gribanova I., et al.* Encoding cryptographic functions to SAT using TRANSALG system // ECAI 2016 — 22nd European Conference on Artificial Intelligence. Frontiers in Artificial Intelligence and Applications. 2016. V. 285. P. 1594–1595.
9. *Biere A.* Lingeling essentials. A tutorial on design and implementation aspects of the the SAT solver lingeling // Proc. Fifth Pragmatics of SAT Workshop. 2014. V. 27. P. 88.
10. `http://hpc.icc.ru` — Иркутский суперкомпьютерный центр СО РАН. Иркутск: ИДСТУ СО РАН.

УДК 519.14+519.25 DOI 10.17223/2226308X/10/62

РАНЖИРОВАНИЕ ПОКАЗАТЕЛЕЙ, ФОРМИРУЮЩИХ КЛАСТЕРНОЕ РАЗБИЕНИЕ, НА ОСНОВЕ КОЭФФИЦИЕНТОВ ОТНОСИТЕЛЬНОГО СХОДСТВА

С. В. Дронов, Е. А. Евдокимов

Рассматривается задача установления относительной информационной ценности числовых показателей, по близости значений которых производится разбиение конечного множества объектов на кластеры. Вводится коэффициент для оценки относительной силы влияния на вид кластерного разбиения каждого из показателей по сравнению с одним или произвольной совокупностью остальных, а также два коэффициента, позволяющих с разных сторон оценить степень связи двух показателей по отношению к этой структуре (кластерная связь). Предложен новый алгоритм сокращения размерности данных на основе этих коэффициентов, в наибольшей степени оставляющий неизменной кластерную структуру исходного множества объектов. Степень искажения оценивается с использованием кластерной метрики, ранее предложенной одним из авторов. Путём реализации этого алгоритма может быть достигнуто более уверенное распознавание угроз компьютерной безопасности при общем снижении нагрузки на систему.

Ключевые слова: *кластерное разбиение, сокращение размерности, кластерная связь, коэффициент силы связи.*

Рассмотрим задачу разбиения конечного множества объектов на кластеры по степени близости совокупностей показателей, которые в этом контексте будем называть формирующими. Нас будет интересовать только результат разбиения, причём договоримся считать, что по совокупности всех рассматриваемых показателей кластеризация объектов производится абсолютно правильно. Мы хотим определить сравнительную силу формирующих показателей по степени их влияния на кластеры. Кроме этого, некоторые из показателей могут быть схожи между собой до такой степени, что использование их вместе совсем не требуется. Такую схожесть показателей для кластерного анализа данных назовём кластерной связью. Силу этой связи тоже можно оценивать с помощью определённых числовых коэффициентов.

Подобные разновидности задачи сокращения размерности данных, по сути являющиеся вариантами post-hoc анализа кластерных разбиений, могут находить примене-