Секция 2

ДИСКРЕТНЫЕ ФУНКЦИИ

УДК 519.7 DOI 10.17223/2226308X/15/5

ПРИМЕНЕНИЕ ЭВРИСТИЧЕСКИХ МЕТОДОВ ДЛЯ ПОИСКА БУЛЕВЫХ ФУНКЦИЙ С КРИПТОГРАФИЧЕСКИМИ ХАРАКТЕРИСТИКАМИ[1]

Н. Д. Атутова

Для успешного противостояния шифров линейному и алгебраическому криптоанализу в их структуре необходимо использовать функции с высокой нелинейностью и алгебраической иммунностью. Работа является продолжением исследования, в котором используется комбинированный подход к поиску криптографических булевых функций на основе эвристических методов, в частности генетического алгоритма и поиска восхождением к вершине (алгоритм Hill Climbing). Проведён сравнительный анализ вариаций мутации и скрещивания для генетического алгоритма и сравнительный анализ ранее полученных результатов со случайным поиском. На основе полученных булевых функций построены векторные булевы функции и среди них подсчитано количество функций, обаладающих высокой компонентной алгебраической иммунностью и нелинейностью.

Ключевые слова: *генетический алгоритм, алгоритм Hill Climbing, алгебраическая иммунность, нелинейность, эвристики.*

Успех криптоанализа в основном зависит от нахождения уязвимостей в математических свойствах булевых функций, являющихся компонентами шифра. Для противостояния разным видам криптоанализа существуют различные математические требования к функциям.

Функции, которые удовлетворяют этим требованиям, называются криптографическими. Для повышения стойкости необходимо искать и применять в шифрах булевы функции с оптимальными криптографическими свойствами. Для ослабления статистической зависимости между входом и выходом функции по умолчанию предполагается, что функция обладает сбалансированностью, т. е. принимает значения 0 и 1 одинаково часто.

Работа является продолжением исследования, начальные результаты которого представлены в [1]. Рассматриваются булевы функции от n переменных с такими криптографическими характеристиками: сбалансированность, нелинейность и алгебраическая иммунность. Известны теоретические оценки максимума этих характеристик: для нелинейности — $N_f \leqslant 2^{n-1} - 2^{n/2-1}$, для алгебраической иммунности — $\mathrm{AI}(f) \leqslant \lceil n/2 \rceil$.

Цель данной работы — построение сбалансированных булевых функций с хорошими криптографическими характеристиками, а также построение на их основе векторных булевых функций с компонентной алгебраической иммунностью и нелинейностью.

[1]Работа выполнена при поддержке Математического центра в Академгородке, соглашение с Министерством науки и высшего образования РФ № 075-15-2022-281.

Функция вида $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ называется *векторной булевой функцией*. Векторную булеву функцию можно представить в виде $F = (f^1(x) \dots f^m(x))$, где $f^j(x)$, $j = 1, \dots, m$, — координатные (булевы) функции от n переменных.

Компонентной функцией называется любая ненулевая линейная комбинация координатных функций, т. е. булева функция $vF = v_1 f_1 \oplus \dots \oplus v_m f_m$, где $v = (v_1 \dots v_m) \in \in \mathbb{F}_2^m \setminus 0^m$ и 0^m — нулевой вектор длины m.

Нелинейностью N_f и *компонентной алгебраической иммунностью* $\mathrm{AI}_{\mathrm{comp}}(F)$ векторной булевой функции $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ называются минимальная нелинейность и алгебраическая иммунность её компонентных функций соответственно:

$$N_F = \min_{v \in \mathbb{F}_2^m \setminus 0^m} N_{vF}, \qquad \mathrm{AI}_{\mathrm{comp}}(F) = \min_{v \in \mathbb{F}_2^m \setminus 0^m} \mathrm{AI}(vF).$$

Можно выделить три способа нахождения таких функций: полный перебор, алгебраическое конструирование и эвристики. Для булевой функции $f : \mathbb{F}_2^n \to \mathbb{F}_2$ длина вектора значений — 2^n. При росте числа переменных количество булевых функций растёт дважды экспоненциально, что делает невозможным полный перебор. Алгебраическое построение заведомо сужает множество решений и повышает вероятность успеха криптоанализа из-за предсказуемости построения. Перспективным является подход, использующий эвристические методы, в основе которых лежит итерационный перебор с параметрами для достижения желаемого результата.

Для поиска криптографических булевых функций и построения на их основе векторных булевых функций предложен гибридный алгоритм на основе генетического алгоритма и поиска восхождением к вершине (алгоритм Hill Climbing) и проведены вычислительные эксперименты, показывающие его эффективность.

Генетический алгоритм (ГА) — это эвристический подход к комбинаторной оптимизации в сложном пространстве пригодности. Идея, лежащая в основе генетического алгоритма, состоит в том, чтобы начать с набора потенциальных решений (генофонда или родительского множества) и каким-то образом объединить их пары (схема размножения) для получения новых решений (дочерних).

Схемы размножения могут включать родительский отбор и мутацию полученного в результате ребенка. Новый набор выбирается из числа родителей и детей (отбор) на основе функции пригодности, и процесс повторяется до тех пор, пока не будет найдено оптимальное решение или не будет достигнуто максимальное число итераций.

В работе особями являются векторы значений булевых функций от n переменных, начальная популяция — случайное множество сбалансированных особей.

Поиск восхождением к вершине (Hill Climbing) — эвристический алгоритм, который начинает с произвольного решения, а затем итерационно пытается найти лучшее решение путём пошагового изменения одного из его элементов. На вход поступает вектор значений булевой функции. Алгоритм итеративно пытается его улучшить, изменяя одну из координат.

Более подробное описание алгоритма и теоретическое обоснование его эффективности представлено в [2]. В работе поиск восхождением к вершине применяется после оператора мутации потомков на каждой итерации ГА для поддержания нелинейности.

Рассмотрим два варианта реализации оператора скрещивания.

Первый — однородное скрещивание. На вход поступают булевы функции f и g от n переменных, заданные векторами значений $(f_0, f_1, \dots, f_{2^n-1})$ и $(g_0, g_1, \dots, g_{2^n-1})$ соответственно. На выходе булева функция h от n переменных, вектор значений $(h_0, h_1, \dots, h_{2^n-1})$ которой определяется следующим образом: если $f_i = g_i$, то $h_i = f_i$;

если $f_i \neq g_i$, то h_i принимается равным значению f_i или g_i с одинаковой вероятностью, где $i = 0, 1, \ldots, 2^{n-1}$.

Второй вариант предложен в [3], он похож на предыдущий, но, в отличие от первого, позволяет поддерживать сбалансированность потомков.

Введём некоторые ограничения на скрещивание. *Расстоянием Хэмминга* ($\mathrm{dist}(f, g)$) между булевыми функциями f и g от n переменных называется число координат, в которых различаются их векторы значений.

Если $\mathrm{dist}(f, g) > 2^{n-1}$, то вместо вектора значений функции g рассматривается вектор, полученный инверсией всех битов вектора значений функции g.

Рассмотрено применение двух видов мутации.

Первый — к двум различным случайным битам входного вектора значений применялась инверсия.

Второй предложен в [4]. Выбираются две случайные позиции в векторе значений. Значение из большей позиции переставляется в меньшую и происходит сдвиг битов вправо.

Для генетического алгоритма с целевой функцией $\mathrm{cost} = \mathrm{AI}(f)$ проведён сравнительный анализ эффективности различных вариантов операций скрещивания и мутации. Полученные результаты приведены в табл. 1, где n — число переменных; P — размер популяции; T — число итераций.

Таблица 1

Результаты применения ГА с различными операторами

n	P	T	max, min, среднее значение $\mathrm{AI}(f)$ в исходной популяции	max, min, среднее значение $\mathrm{AI}(f)$ после применения ГА	Количество функций с $\max \mathrm{AI}(f)$	Количество функций с $\max \mathrm{AI}(f)$ с новыми операторами
6	10	20	(3, 1, 1,9)	(3, 3, 3)	659	682
7	10	20	(4, 2, 2,8)	(4, 4, 4)	14	34
8	10	20	(4, 1, 2,7)	(4, 4, 4)	671	831

Как видно из табл. 1, все потомки имеют ненулевую алгебраическую иммунность вне зависимости от вариации операторов, но неклассические мутации находят большее число функций с максимальной алгебраической иммунностью.

Для оценки эффективности гибридного подхода, основанного на сочетании генетического алгоритма с новыми операторами и поиска восхождением к вершине проведён сравнительный анализ по времени и количеству найденных функций со случайным перебором из одного миллиона случайно сгенерированных сбалансированных функций. Результаты представлены в табл. 2.

Таблица 2

Сравнение гибридного подхода со случайным поиском

n	P	T	Найдено функций ГА+НС	Время, с	Случайный поиск из 1000000 функций	Время, с
4	20	5000	314222	432	91002	972
6	20	5000	564111	765	123611	1187
7	20	5000	457211	1056	198715	1623
8	20	5000	732888	1434	261238	2111

Как видно из табл. 2, гибридный подход показывает высокую эффективность в задаче поиска булевых функций с алгебраической иммунностью ($\mathrm{cost} = \mathrm{AI}(f)$).

Полученные булевы функции используются для построения векторных булевых функций. Изменим целевую функцию на сочетание криптографических характеристик и применим гибридный подход для поиска булевых функций и на основе получившихся построим векторные булевы функции для $m = 4$ и $n = 4, 5, 6, 8$. Целевая функция — нелинейность и алгебраическая иммунность: $\text{cost} = \dfrac{N_f}{\max N_g} + \dfrac{\text{AI}(f)}{\max \text{AI}(g)}$, где g пробегает множество особей популяции. Результаты представлениы в табл. 3

Таблица 3

Результаты для векторных булевых функций

n	P	T	Найдено булевых функций	Найдено векторных булевых функций	Значение N_F и $\text{AI}_{\text{comp}}(F)$	Теор. оценка N_F	Теор. оценка $\text{AI}_{\text{comp}}(F)$	Время, с
4	20	10	1245	415661	6; 2	$\leqslant 6$	$\leqslant 2$	162,202
5	20	10	76	1797608	12; 3	< 13	$\leqslant 3$	5323,85
6	20	10	1346	310295	24; 3	$\leqslant 28$	$\leqslant 3$	363,098
8	20	10	1426	321203	108; 4	$\leqslant 120$	$\leqslant 4$	3061,11

Таким образом, в работе приведены результаты сравнительного анализа гибридного подхода и случайного перебора. Найденные булевы функции использованы для поиска векторных булевых функций. Представлены результаты поиска векторных булевых функций с нелинейностью и компонентной алгебраической иммунностью для $m = 4$ и $n \leqslant 8$. Полученные векторные булевы функции могут быть использованы для построения S-блоков. Наличие компонентной алгебраической иммунности и нелинейности S-блоков способствует противостоянию алгебраическому и линейному криптоанализу шифров соответственно.

ЛИТЕРАТУРА

1. *Амутова Н. Д.* Гибридный подход к поиску булевых функций с высокой алгебраической иммунностью на основе эвристических методов // Прикладная дискретная математика. Приложение. 2021. № 14. С. 37–40.
2. *Millan W., Clark A., and Dawson E.* An effective genetic algorithm for finding highly nonlinear Boolean functions // LNCS. 1997. V. 1334. P. 149–158.
3. *Kang M. and Wang M.* New genetic operators for developing S-boxes with low boomerang uniformity // IEEE Access. 2022. V. 10. P. 10898–10906.
4. *Behera P. and Gangopadhyay S.* Evolving bijective S-Boxes using hybrid adaptive genetic algorithm with optimal cryptographic properties // J. Ambient Intell. Human. Comput. 2021. P. 2640–2658.