МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ПРИКЛАДНАЯ ДИСКРЕТНАЯ МАТЕМАТИКА

Научный журнал

2018 № 41

Зарегистрирован в Федеральной службе по надзору
в сфере связи и массовых коммуникаций

Свидетельство о регистрации ПИ № ФС 77-33762 от 16 октября 2008 г.

Подписной индекс в объединённом каталоге «Пресса России» 38696

УЧРЕДИТЕЛЬ
Томский государственный университет

Адрес редакции и издателя: 634050, г. Томск, пр. Ленина, 36
E-mail: vestnik_pdm@mail.tsu.ru

В журнале публикуются результаты фундаментальных и прикладных научных исследований отечественных и зарубежных ученых, включая студентов и аспирантов, в области дискретной математики и её приложений в криптографии, компьютерной безопасности, кибернетике, информатике, программировании, теории надёжности, интеллектуальных системах.

Периодичность выхода журнала: 4 номера в год.

Редактор *Н. И. Шидловская*
Верстка *И. А. Панкратовой*

Подписано к печати 17.09.2018. Формат $60 \times 84\frac{1}{8}$. Усл. п. л. 14,6. Тираж 300 экз.
Заказ № 3366. Цена свободная. Дата выхода в свет 28.09.2018.

Отпечатано на оборудовании
Издательского Дома Томского государственного университета
634050, г. Томск, пр. Ленина, 36
Тел.: 8(3822)53-15-28, 52-98-49

СОДЕРЖАНИЕ

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

МАТЕМАТИЧЕСКИЕ МЕТОДЫ КРИПТОГРАФИИ

ПРИКЛАДНАЯ ТЕОРИЯ ГРАФОВ

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ И ПРОГРАММИРОВАНИЯ

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ В ДИСКРЕТНОЙ МАТЕМАТИКЕ

CONTENTS

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

УДК 519.115, 519.113.5, 512.624, 512.552.7

О ЧИСЛЕ ОДНОРОДНЫХ НЕВЫРОЖДЕННЫХ p-ИЧНЫХ ФУНКЦИЙ ЗАДАННОЙ СТЕПЕНИ[1]

М. И. Анохин

Институт проблем информационной безопасности Московского государственного университета имени М. В. Ломоносова, г. Москва, Россия

Пусть p — простое число, $F = \mathrm{GF}(p)$, V_n — n-мерное векторное пространство над F, e — базис пространства V_n. Пусть также $\varphi\colon V_n \to F$. Функция φ называется e-однородной, если $\varphi(x) = \pi_{\varphi,e}(\mathbf{x})$ для всех $x \in V_n$, где $\pi_{\varphi,e}$ — однородный многочлен от n переменных над F, имеющий степень не более $p-1$ по каждой переменной, а $\mathbf{x}$ — набор координат вектора x в базисе e. Функция φ называется невырожденной, если $\deg\varphi \geqslant 1$ и $\deg \partial_v\varphi = (\deg\varphi) - 1$ для любого $v \in V_n \setminus \{0\}$, где $(\partial_v\varphi)(x) = \varphi(x+v) - \varphi(x)$ для всех $v, x \in V_n$. Получена формула для числа $\mathrm{HN}_p(n,d)$ e-однородных невырожденных функций $\varphi\colon V_n \to F$, имеющих степень d (это число не зависит от e), а именно: если $n \geqslant 1$ и $d \in \{1,\dots,n(p-1)\}$, то $\mathrm{HN}_p(n,d) = \sum\limits_{k=0}^{n} (-1)^k p^{\binom{k}{2}+\left\{{n-k \atop d}\right\}_p} \begin{bmatrix} n \\ k \end{bmatrix}_p = \sum\limits_{S\subseteq\{1,\dots,n\}} (-1)^{|S|} p^{\sigma(S)-|S|+\left\{{n-|S| \atop d}\right\}_p}$, где $\left\{{m \atop d}\right\}_p$ — обобщённый биномиальный коэффициент порядка p; $\begin{bmatrix} n \\ k \end{bmatrix}_p$ — биномиальный коэффициент Гаусса; $\sigma(S)$ — сумма всех элементов множества S. Доказано, что $\mathrm{HN}_p(n,d) \geqslant p^{\left\{{n \atop d}\right\}_p} - 1 - (p^n-1)\left(p^{\left\{{n-1 \atop d}\right\}_p} - 1\right)/(p-1)$ для любых $d \geqslant 1$ и $n \geqslant d/(p-1)$. Используя эту оценку, получаем, что если $d \geqslant 3$, то $\mathrm{HN}_p(n,d) \sim p^{\left\{{n \atop d}\right\}_p}$ при $n \to \infty$.

Ключевые слова: *p-ичная функция, однородная функция, невырожденная функция, степень функции, формула обращения Мёбиуса, групповая алгебра, фундаментальный идеал, базис Дженнингса.*

DOI 10.17223/20710410/41/1

ON THE NUMBER OF HOMOGENEOUS NONDEGENERATE p-ARY FUNCTIONS OF THE GIVEN DEGREE

M. I. Anokhin

Information Security Institute, Lomonosov University, Moscow, Russia

E-mail: anokhin@mccme.ru

[1]Работа поддержана грантом РФФИ № 16-01-00226.

Let p be a prime number and $F = \mathrm{GF}(p)$. Suppose V_n is an n-dimensional vector space over F and e is a basis of V_n. Also, let $\varphi\colon V_n \to F$. The function φ is called e-homogeneous if $\varphi(x) = \pi_{\varphi,e}(\mathbf{x})$ for all $x \in V_n$, where $\pi_{\varphi,e}$ is an n-variate homogeneous polynomial over F of degree at most $p-1$ in each variable and $\mathbf{x}$ is the coordinate vector of x with respect to the basis e. The function φ is said to be nondegenerate if $\deg\varphi \geqslant 1$ and $\deg\partial_v\varphi = (\deg\varphi) - 1$ for any $v \in V_n \setminus \{0\}$, where $(\partial_v\varphi)(x) = \varphi(x+v) - \varphi(x)$ for all $v, x \in V_n$. This notion was introduced by O. A. Logachev, A. A. Sal'nikov, and V. V. Yashchenko in the case when $p = 2$. Our main results are as follows. First, we obtain a formula for the number $\mathrm{HN}_p(n,d)$ of e-homogeneous nondegenerate functions $\varphi\colon V_n \to F$ of degree d (this number does not depend on e). Namely, if $n \geqslant 1$ and $d \in \{1,\ldots,n(p-1)\}$, then $\mathrm{HN}_p(n,d) = \sum\limits_{k=0}^{n}(-1)^k p^{\binom{k}{2}+\left\{{n-k \atop d}\right\}_p}\begin{bmatrix} n \\ k \end{bmatrix}_p = \sum\limits_{S\subseteq\{1,\ldots,n\}}(-1)^{|S|}p^{\sigma(S)-|S|+\left\{{n-|S| \atop d}\right\}_p}$, where $\left\{{m \atop d}\right\}_p$ is the generalized binomial coefficient of order p, $\begin{bmatrix} n \\ k \end{bmatrix}_p$ is the Gaussian binomial coefficient, and $\sigma(S)$ is the sum of all elements of S. The proof of this formula is based on the Möbius inversion. Previously, only formulas for $\mathrm{HN}_p(n,2)$ were known; unlike our formula, their forms depend on the parities of p and n. Second, we prove that $\mathrm{HN}_p(n,d) \geqslant p^{\left\{{n \atop d}\right\}_p - 1 - (p^n-1)\left(p^{\left\{{n-1 \atop d}\right\}_p}-1\right)/(p-1)}$ for any $d \geqslant 1$ and $n \geqslant d/(p-1)$. Using this bound, we obtain that if $d \geqslant 3$, then $\mathrm{HN}_p(n,d) \sim p^{\left\{{n \atop d}\right\}_p}$ as $n \to \infty$. For $p = 2$ the last two statements were proved by Yu. V. Kuznetsov. The proofs of our main results use a Jennings basis of the group algebra FG_n, where G_n is an elementary abelian p-group of rank n.

Keywords: *p-ary function, homogeneous function, nondegenerate function, degree of a function, Möbius inversion formula, group algebra, augmentation ideal, Jennings basis.*

1. Определения, обозначения и необходимые факты

Фиксируем простое число p и обозначим через F поле из p элементов. В настоящей работе все объекты и понятия линейной алгебры рассматриваются над основным полем F. Для произвольного множества X через F^X обозначается множество всех функций из X в F. Это множество является векторным пространством относительно поточечных операций сложения и умножения на элементы из F. Если S — какая-либо система векторов произвольного векторного пространства, то $\langle S\rangle$ обозначает линейную оболочку этой системы. Пусть также n — произвольное целое неотрицательное число и V_n — n-мерное векторное пространство (над полем F). Элементы множества F^{V_n} естественно назвать *p-ичными функциями*.

Выберем какой-либо базис $e = (e_1,\ldots,e_n)$ пространства V_n. Иногда будем выбирать этот базис некоторым специальным образом. Пусть также $\varphi \in F^{V_n}$. Тогда хорошо известно, что существует единственный многочлен $\pi_{\varphi,e} = \pi_{\varphi,e}(t) \in F[t_1,\ldots,t_n]$ (зависящий от φ и e), который имеет степень не более $p-1$ по каждой переменной и удовлетворяет равенству

$$\varphi(x_1e_1 + \ldots + x_ne_n) = \pi_{\varphi,e}(x_1,\ldots,x_n)$$

для любых $x_1,\ldots,x_n \in F$. Другими словами, если x — произвольный вектор из V_n и $\mathbf{x}$ — набор координат этого вектора в базисе e, то $\varphi(x) = \pi_{\varphi,e}(\mathbf{x})$. Функция $\varphi \mapsto \pi_{\varphi,e}$

($\varphi \in F^{V_n}$) является изоморфизмом векторного пространства F^{V_n} на векторное пространство всех многочленов из $F[t_1, \ldots, t_n]$, имеющих степень не более $p-1$ по каждой переменной.

Напомним, что *производной* функции φ по направлению $v \in V_n$ называется функция $\partial_v\varphi \in F^{V_n}$, определённая равенством $(\partial_v\varphi)(x) = \varphi(x+v)-\varphi(x)$ для каждого $x \in V_n$. *Степенью* функции φ, обозначаемой $\deg\varphi$, называется степень многочлена $\pi_{\varphi,e}$. Считаем, что степень нулевого многочлена (а следовательно, и степень нулевой функции из F^{V_n}) равна -1. Определённая таким образом степень функции φ не зависит от выбора базиса e, так как она совпадает с наименьшим целым числом $d \geqslant -1$, для которого $\partial_{v_1} \ldots \partial_{v_{d+1}}\varphi = 0$ при всех $v_1, \ldots, v_{d+1} \in V_n$ [1, свойство B7]. Очевидно, что $-1 \leqslant \deg\varphi \leqslant n(p-1)$ и что если $\deg\varphi \geqslant 0$, то $\deg\partial_v\varphi \leqslant \deg\varphi - 1$ для любого $v \in V_n$. Кроме того, если $d \in \{-1, \ldots, n(p-1)\}$, то множество $\mathrm{RM}_d = \{\psi \in F^{V_n} : \deg\psi \leqslant d\}$ является подпространством пространства F^{V_n}. Обозначение RM_d связано с тем, что именно так можно определить p-ичный код Рида—Маллера порядка d и длины p^n.

Функция φ называется *e-однородной*, если $\pi_{\varphi,e}$ является однородным многочленом (к которым относится и нулевой многочлен). Очевидно, что всякая e-однородная функция степени не более $p-1$ будет также e'-однородной для любого базиса e' пространства V_n. Поэтому можно говорить просто об однородных функциях степени не более $p-1$. Однако в общем случае однородность функции φ зависит от выбора базиса e. Например, пусть $n = 2$ и функция $\psi \in F^{V_2}$ такова, что $\pi_{\psi,e} = t_1^{p-1}t_2$. Тогда если $e' = (e_1 + e_2, e_2)$, то $\pi_{\psi,e'} = t_1 + t_1^{p-1}t_2$. Таким образом, функция ψ e-однородна, но не e'-однородна.

Замечание 1. Очевидно, что функция φ представима единственным образом в виде $\varphi_{(0,e)} + \ldots + \varphi_{(d,e)}$, где $d = \deg\varphi$, а $\varphi_{(i,e)}$ — либо e-однородная функция степени i из F^{V_n}, либо нулевая функция из F^{V_n} ($i \in \{0, \ldots, d\}$), причём если $d \geqslant 0$, то $\varphi_{(d,e)} \neq 0$. Функция $\varphi_{(i,e)}$ называется *i-й e-однородной компонентой* функции φ.

Предположим, что $\deg\varphi \geqslant 1$ (это возможно, если и только если $n \geqslant 1$). Тогда полагаем

$$L(\varphi) = \{v \in V_n : \deg\partial_v\varphi \leqslant \deg\varphi - 2\}.$$

Легко видеть, что $L(\varphi)$ является собственным подпространством пространства V_n. Можно также сказать, что $L(\varphi)$ — группа инерции смежного класса $\varphi + \mathrm{RM}_{(\deg\varphi)-2}$ в группе сдвигов пространства V_n, если отождествить сдвиг $x \mapsto x + v$ на вектор $v \in V_n$ с самим этим вектором. Очевидно, что RM_d для любого $d \in \{-1, \ldots, n(p-1)\}$ замкнуто относительно сдвигов аргумента функции. Будем пользоваться тем, что если $\varphi' \in \varphi + \mathrm{RM}_{(\deg\varphi)-1}$, то $L(\varphi') = L(\varphi)$. Функция φ называется *невырожденной*, если $L(\varphi) = \{0\}$, или, что эквивалентно, $\deg\partial_v\varphi = \deg\varphi - 1$ для любого $v \in V_n \setminus \{0\}$. Понятие невырожденной функции введено в [2] в случае, когда $p = 2$ (см. также [3, определение 3.5.2]). Основной смысл этого понятия состоит в следующей теореме (в случае $p = 2$ см. [2, теорема 3] или [3, теорема 3.5.3]):

Теорема 1. Пусть $\varphi \in F^{V_n}$, причём $\deg\varphi = d \geqslant 1$. Предположим, что $L(\varphi) = \langle e_{m+1}, \ldots, e_n\rangle$ для некоторого $m \in \{1, \ldots, n\}$. Тогда существуют функции $\varphi' \in F^{V_n}$ и $\psi \in F^{V_n/L(\varphi)}$, удовлетворяющие следующим условиям:

1) $\deg\varphi' \leqslant d - 1$;
2) ψ имеет степень d, $(e_1 + L(\varphi), \ldots, e_m + L(\varphi))$-однородна и невырождена;
3) $\varphi(x) = \varphi'(x) + \psi(x + L(\varphi))$ для всех $x \in V_n$.

Теорема 1 может быть легко доказана по схеме, описанной в [2] и [3, разд. 3.5] (хотя в обоих этих источниках подробное доказательство не приводится), а именно: пусть

$\varphi_{(d,e)}$ — d-я e-однородная компонента функции φ (см. замечание 1). Непосредственно проверяется, что многочлен $\pi_{\varphi_{(d,e)},e}$ не зависит от переменных $t_{m+1},\dots,t_n$. Следовательно, существует функция $\psi \in F^{V_n/L(\varphi)}$, такая, что $\varphi_{(d,e)}(x) = \psi(x + L(\varphi))$ для всех $x \in V_n$. Из равенства

$$\pi_{\psi,(e_1+L(\varphi),\dots,e_m+L(\varphi))}(t_1,\dots,t_m) = \pi_{\varphi_{(d,e)},e}(t_1,\dots,t_m)$$

вытекает п. 2 теоремы 1. Поэтому $\varphi' = \varphi - \varphi_{(d,e)}$ и ψ удовлетворяют условиям этой теоремы.

Пусть $d \in \{1,\dots,n(p-1)\}$ (и, следовательно, $n \geqslant 1$). В связи с теоремой 1 представляет интерес нахождение числа e-однородных невырожденных функций из F^{V_n}, имеющих степень d. Это число будем обозначать через $\mathrm{HN}_p(n,d)$.

Замечание 2. Пусть $\alpha \in (\mathrm{RM}_d/\mathrm{RM}_{d-1}) \setminus \{0\}$, где $d \in \{1,\dots,n(p-1)\}$. Определим $L(\alpha)$ как $L(\chi)$ для произвольной функции χ из смежного класса α; корректность этого определения очевидна. Из замечания 1 следует, что в смежном классе α содержится ровно одна e-однородная функция, а именно d-я e-однородная компонента произвольной функции из этого смежного класса. Это показывает, что

$$\mathrm{HN}_p(n,d) = |\{\alpha \in (\mathrm{RM}_d/\mathrm{RM}_{d-1}) \setminus \{0\} : L(\alpha) = \{0\}\}|.$$

В частности, $\mathrm{HN}_p(n,d)$ не зависит от выбора базиса e.

Легко видеть, что

$$\mathrm{HN}_p(n,1) = \begin{cases} p-1, & \text{если } n = 1, \\ 0, & \text{если } n \geqslant 2. \end{cases} \tag{1}$$

Приведём теперь формулы для $\mathrm{HN}_p(n,2)$. Пусть φ — e-однородная функция степени 2 из F^{V_n}. Тогда существует единственная ненулевая матрица $A_{\varphi,e}$ размера $n \times n$ над F с нулями под главной диагональю (а при $p = 2$ и на главной диагонали), такая, что $\pi_{\varphi,e} = tA_{\varphi,e}t^{\mathrm{T}}$. Здесь и далее $t = (t_1,\dots,t_n)$ — набор переменных, а верхний индекс «$^{\mathrm{T}}$» обозначает транспонирование. Непосредственно проверяется, что $\pi_{\partial_v\varphi,e} = \mathbf{v}(A_{\varphi,e} + A_{\varphi,e}^{\mathrm{T}})t^{\mathrm{T}} + \mathbf{v}A_{\varphi,e}\mathbf{v}^{\mathrm{T}}$ для любого $v \in V_n$, где $\mathbf{v}$ — набор координат вектора v в базисе e. Поэтому функция φ невырождена тогда и только тогда, когда матрица $A_{\varphi,e} + A_{\varphi,e}^{\mathrm{T}}$ невырождена. Отметим, что $A_{\varphi,e} + A_{\varphi,e}^{\mathrm{T}}$ является матрицей симметричной билинейной формы $(x,y) \mapsto \varphi(x+y) - \varphi(x) - \varphi(y)$ $(x,y \in V_n)$, ассоциированной с φ как квадратичной формой, в базисе e. Это показывает, что функция φ невырождена в смысле настоящей работы тогда и только тогда, когда она невырождена как квадратичная форма (т. е. ассоциированная с ней билинейная форма невырождена).

Из этих рассуждений следует, что функция $\varphi \mapsto A_{\varphi,e} + A_{\varphi,e}^{\mathrm{T}}$ инъективно отображает множество всех e-однородных невырожденных функций степени 2 из F^{V_n} на множество всех невырожденных симметричных матриц размера $n \times n$ над F (а при $p = 2$, кроме того, имеющих на главной диагонали лишь нули). Поэтому $\mathrm{HN}_p(n,2)$ равно числу таких матриц, которое давно известно. Это число можно найти, используя классификацию невырожденных симметричных (при $p \neq 2$) и симплектических (при $p = 2$) билинейных форм над F и формулы для порядков линейных групп, сохраняющих эти формы. Однако удобнее воспользоваться формулами, приведёнными в теоремах 2 и 3 работы [4]. Таким образом, если q — нечётное простое число и m — целое положительное число, то

$$\begin{aligned} &\mathrm{HN}_q(2m-1,2) = q^{m^2-m}\prod_{i=0}^{m-1}(q^{2i+1}-1), \quad \mathrm{HN}_q(2m,2) = q^{m^2+m}\prod_{i=0}^{m-1}(q^{2i+1}-1),\\ &\mathrm{HN}_2(2m-1,2) = 0, \qquad\qquad\qquad\quad \mathrm{HN}_2(2m,2) = 2^{m^2-m}\prod_{i=0}^{m-1}(2^{2i+1}-1). \end{aligned} \tag{2}$$

2. Формулировки основных результатов

Одним из основных результатов настоящей работы является формула для $\mathrm{HN}_p(n,d)$ (см. теорему 2 ниже). Эта формула доказывается с помощью формулы обращения Мёбиуса. В отличие от формул (2) для $\mathrm{HN}_p(n,2)$, вид нашей формулы не зависит ни от чётности p, ни от чётности n. Введём некоторые обозначения. Пусть m — целое неотрицательное число. Положим

$$J_m = \{(j_1,\dots,j_m) : j_1,\dots,j_m \in \{0,\dots,p-1\}\}. \tag{3}$$

Для каждого $j \in J_m$ обозначим через $\sigma(j)$ сумму всех элементов набора j. Тогда для произвольного целого числа k *обобщённый биномиальный коэффициент* $\left\{{m \atop k}\right\}_p$ *порядка* p определяется как $|\{j \in J_m : \sigma(j) = k\}|$. Отметим, что при $k \geqslant 0$ $\left\{{m \atop k}\right\}_p$ — это число способов размещения k одинаковых предметов в m ячейках, в каждой из которых число предметов не может превосходить $p-1$ [5, разд. 1.3]. Если $k \leqslant -1$, то $\left\{{m \atop k}\right\}_p = 0$. Очевидно, что функция $(j_1,\dots,j_m) \mapsto (p-1-j_1,\dots,p-1-j_m)$ инъективно отображает множество $\{j \in J_m : \sigma(j) = k\}$ на множество $\{j \in J_m : \sigma(j) = m(p-1)-k\}$. Поэтому

$$\left\{{m \atop m(p-1)-k}\right\}_p = \left\{{m \atop k}\right\}_p \tag{4}$$

для любого целого числа k (см. также формулу (1.14) из [5]).

Кроме того, для произвольного $k \in \{0,\dots,n\}$ через $\begin{bmatrix} n \\ k \end{bmatrix}_p$ будем обозначать число k-мерных подпространств пространства V_n (*биномиальный коэффициент Гаусса*, или *квантовый биномиальный коэффициент*). Известно, что

$$\begin{bmatrix} n \\ k \end{bmatrix}_p = \prod_{i=1}^{k}\frac{p^{n-i+1}-1}{p^i-1} = \frac{\prod_{i=1}^{n}(p^i-1)}{\prod_{i=1}^{k}(p^i-1)\prod_{i=1}^{n-k}(p^i-1)} \tag{5}$$

(см. пример 2.64 и подразд. C разд. 1 гл. III (в частности, формулы (3.11) и (3.12) из [6], а также формулу (6.4) из [7] и формулы (1.5), (1.6) из [5]).

Теорема 2. Пусть $n \geqslant 1$ и $d \in \{1,\dots,n(p-1)\}$. Тогда

$$\mathrm{HN}_p(n,d) = \sum_{k=0}^{n}(-1)^k p^{\binom{k}{2}+\left\{{n-k \atop d}\right\}_p}\begin{bmatrix} n \\ k \end{bmatrix}_p = \sum_{S\subseteq\{1,\dots,n\}}(-1)^{|S|}p^{\sigma(S)-|S|+\left\{{n-|S| \atop d}\right\}_p}, \tag{6}$$

где $\sigma(S)$ — сумма всех элементов множества $S \subseteq \{1,\dots,n\}$.

Формула (6) имеет не очень простой вид. Поэтому представляет интерес вопрос об асимптотике $\mathrm{HN}_p(n,d)$ при фиксированных p и d и при $n\to\infty$. Случай, когда $d=1$, тривиален ввиду формулы (1). Рассмотрим теперь случай, когда $d=2$. Пусть, как и в формулах (2), q — нечётное простое число и m — целое положительное число. Нам потребуется понятие *квантового символа Похгаммера*: если $0<a,b<1$, то

$$(a;b)_m=\prod_{i=0}^{m-1}(1-ab^i)\quad\text{и}\quad(a;b)_\infty=\lim_{m\to\infty}(a;b)_m=\prod_{i=0}^{\infty}(1-ab^i).$$

Легко видеть, что бесконечное произведение в определении $(a;b)_\infty$ сходится, причём $0<(a;b)_\infty<1$. В частности,

$$(1/p;1/p^2)_m=\prod_{i=0}^{m-1}\left(1-\frac{1}{p^{2i+1}}\right)\quad\text{и}\quad(1/p;1/p^2)_\infty=\prod_{i=0}^{\infty}\left(1-\frac{1}{p^{2i+1}}\right).$$

Из известного равенства $\sum\limits_{i=0}^{m-1}(2i+1)=m^2$ следует, что $\prod\limits_{i=0}^{m-1}(p^{2i+1}-1)=p^{m^2}(1/p;1/p^2)_m$. Используя последнее равенство и формулы (2), получаем

$$\mathrm{HN}_q(2m-1,2)=q^{\binom{2m}{2}}(1/q;1/q^2)_m,\quad \mathrm{HN}_q(2m,2)=q^{\binom{2m+1}{2}}(1/q;1/q^2)_m,$$
$$\mathrm{HN}_2(2m,2)=2^{\binom{2m}{2}}(1/2;1/4)_m.$$

Следовательно,

$$\mathrm{HN}_q(n,2)\sim q^{\binom{n+1}{2}}(1/q;1/q^2)_\infty\quad\text{и}\quad\mathrm{HN}_2(2m,2)\sim 2^{\binom{2m}{2}}(1/2;1/4)_\infty \tag{7}$$

при $n\to\infty$ и $m\to\infty$ соответственно. Здесь и далее $\sim$ обозначает асимптотическую эквивалентность функций при стремлении к ∞ некоторого указанного целого неотрицательного аргумента этих функций; при этом все остальные аргументы и параметры функций, если они есть, предполагаются фиксированными.

Перейдём к случаю, когда $d\geqslant 3$.

Теорема 3. Пусть $d\geqslant 1$ и $n\geqslant d/(p-1)$. Тогда

$$\mathrm{HN}_p(n,d)\geqslant p^{\left\{{n\atop d}\right\}_p}-1-\frac{p^n-1}{p-1}\left(p^{\left\{{n-1\atop d}\right\}_p}-1\right). \tag{8}$$

Кроме того,

$$d\geqslant 3\implies \mathrm{HN}_p(n,d)\sim p^{\left\{{n\atop d}\right\}_p}\text{ при }n\to\infty. \tag{9}$$

В случае $p=2$ теорема 3 доказана Ю. В. Кузнецовым в [8]. Доказательство теоремы 3 проводится по той же схеме, что и доказательство её частного случая в [8].

Пусть $d\in\{-1,\dots,n(p-1)\}$. Обозначим через $\mathrm{H}_p(n,d)$ число e-однородных функций из F^{V_n}, имеющих степень d (как мы увидим ниже, это число не зависит от выбора базиса e). Легко видеть, что множество, состоящее из нулевой функции из F^{V_n} и всех e-однородных функций степени d из F^{V_n}, является $\left\{{n\atop d}\right\}_p$-мерным подпространством пространства F^{V_n}. Следовательно,

$$\mathrm{H}_p(n,d)=\begin{cases}1, & \text{если } d=-1,\\ p^{\left\{{n\atop d}\right\}_p}-1, & \text{если } d\geqslant 0.\end{cases} \tag{10}$$

В частности, $\mathrm{H}_p(n,0)=p-1$.

Предположим теперь, что $d \geqslant 1$. Тогда легко видеть, что $\left\{ {n \atop d} \right\}_p \to +\infty$ при $n \to \infty$, так как $\left\{ {n \atop d} \right\}_p \geqslant \binom{n}{d} \sim n^d/d!$ при $n \to \infty$. Поэтому

$$d \geqslant 1 \implies \mathrm{H}_p(n,d) \sim p^{\left\{ {n \atop d} \right\}_p} \text{ при } n \to \infty \tag{11}$$

ввиду формулы (10). Это вместе с некоторыми приведёнными выше формулами позволяет найти предел доли невырожденных функций среди e-однородных функций степени d из F^{V_n} (т. е. предел $\mathrm{HN}_p(n,d)/\mathrm{H}_p(n,d)$) при $n \to \infty$, если он существует, а именно: при $d = 2$ и $p = 2$ этого предела не существует, так как $\mathrm{HN}_2(2m-1,2)/\mathrm{H}_2(2m-1,2) = 0$ при всех $m \geqslant 2$ (см. (2) и (10)) и $\mathrm{HN}_2(2m,2)/\mathrm{H}_2(2m,2) \to (1/2;1/4)_\infty > 0$ при $m \to \infty$ ввиду (7), (11) и очевидного равенства $\left\{ {n \atop 2} \right\}_2 = \binom{n}{2}$. Во всех остальных случаях вышеупомянутый предел указан в формуле

$$\lim_{n\to\infty} \frac{\mathrm{HN}_p(n,d)}{\mathrm{H}_p(n,d)} = \begin{cases} 0, & \text{если } d = 1, \\ (1/p;1/p^2)_\infty, & \text{если } d = 2 \text{ и } p \neq 2, \\ 1, & \text{если } d \geqslant 3, \end{cases}$$

которая следует из формул (1), (7), (9) и (11), а также из того, что $\left\{ {n \atop 2} \right\}_p = \binom{n+1}{2}$ при $p \neq 2$.

3. Переход к групповой алгебре элементарной абелевой p-группы над F

С этого момента вместо векторного пространства V_n (в котором умножение на произвольный элемент основного поля F выражается через сложение) будем иметь дело с элементарной абелевой p-группой G_n ранга n. Для этой группы мы используем мультипликативную запись. Разумеется, группа G_n изоморфна аддитивной группе векторного пространства V_n. Поэтому переход от V_n к G_n представляет собой лишь смену обозначений. В частности, все понятия и объекты, определённые в п. 1 и относящиеся к F^{V_n} (например, производная функции, степень функции, пространства RM_d и $L(\varphi)$), имеют естественные аналоги, относящиеся к F^{G_n}. Эти аналоги будут называться и обозначаться так же, как и исходные понятия и объекты. Подпространствам пространства V_n соответствуют подгруппы группы G_n, а размерностям этих подпространств — ранги соответствующих подгрупп.

Множество F^{G_n} является унитарным модулем над групповой алгеброй FG_n, в котором действие FG_n продолжает по линейности действие группы G_n сдвигами аргумента, а именно

$$\left(\left(\sum_{g \in G_n} f_g g \right) \varphi \right)(x) = \sum_{g \in G_n} f_g \varphi(xg)$$

для любых $f_g \in F$ ($g \in G_n$), $\varphi \in F^{G_n}$ и $x \in G_n$. Такой подход используется в [9] (там рассматривается несколько другое действие, однако его отличие от действия настоящей работы не принципиально).

Обозначим через Δ фундаментальный идеал групповой алгебры FG_n, т. е. идеал этой алгебры, порождённый множеством $\{g - 1 : g \in G_n \setminus \{1\}\}$ (легко видеть, что

это множество является также базисом Δ как векторного пространства). Другими словами,

$$\sum_{g\in G_n} f_g g \in \Delta \iff \sum_{g\in G_n} f_g = 0,$$

где $f_g \in F$ $(g \in G_n)$ (см. также теорему 1.2 из [10] или п. (а) леммы 3.1.7 из [11]). Идеал Δ совпадает с радикалом Джекобсона (или, что в данном случае эквивалентно, с наибольшим нильпотентным идеалом) алгебры FG_n (см. теорему 1.2 из [10] или п. (а) теоремы 3.1.9 из [11]). Для целого положительного числа k через Δ^k обозначим k-ю степень идеала Δ; полагаем $\Delta^0 = FG_n$. Известно, что $\Delta^{n(p-1)} \neq \{0\}$, но $\Delta^{n(p-1)+1} = \{0\}$ (см. теоремы 6.2 (или 6.5) и 3.7 из [10], определение и теорему 3.3.12 из [11], а также [12, разд. 1]). Отметим, что производная произвольной функции $\varphi \in F^{G_n}$ по любому направлению $g \in G_n$ может быть записана как $(g-1)\varphi$. Поэтому

$$\mathrm{RM}_d = \{\varphi \in F^{G_n} : \Delta^{d+1}\varphi = \{0\}\} \tag{12}$$

для каждого $d \in \{-1, \ldots, n(p-1)\}$. Из этого, в частности, следует, что RM_d является подмодулем FG_n-модуля F^{G_n}.

Пусть $b = (b_1, \ldots, b_n)$ — базис элементарной абелевой p-группы G_n. Для каждого $j = (j_1, \ldots, j_n) \in J_n$ положим

$$\eta_j(b) = (b_1 - 1)^{j_1} \ldots (b_n - 1)^{j_n}.$$

Здесь множество J_n определено в (3). Напомним также, что если $j \in J_n$, то $\sigma(j)$ — это сумма всех элементов набора j. Тогда для любого целого неотрицательного числа k система $(\eta_j(b) \,|\, j \in J_n,\ \sigma(j) \geqslant k)$ является базисом векторного пространства Δ^k (см. теорему 3.2 из [10], определение и теорему 3.3.12 из [11] или [12, разд. 1]). Этот базис называется *базисом Дженнингса*. В частности,

$$\dim(\Delta^k/\Delta^{k+1}) = \binom{n}{k}_p, \tag{13}$$

так как $(\eta_j(b) + \Delta^{k+1} \,|\, j \in J_n,\ \sigma(j) = k)$ — базис векторного пространства Δ^k/Δ^{k+1} (см. теорему 3.6 из [10], цитируемую в [13] как теорема 3). Очевидно также, что

$$\eta_j(b)\eta_{j'}(b) = \begin{cases} \eta_{j+j'}(b), & \text{если } j + j' \in J_n, \\ 0 & \text{в противном случае} \end{cases} \tag{14}$$

для произвольных $j, j' \in J_n$, где $j + j'$ вычисляется поэлементно. Здесь мы воспользовались тем, что $(g-1)^p = g^p - 1 = 0$ для любого $g \in G_n$.

Легко видеть, что функция

$$\varphi \mapsto \sum_{g\in G_n} \varphi(g)g^{-1} \quad (\varphi \in F^{G_n}) \tag{15}$$

является изоморфизмом F^{G_n} на FG_n как FG_n-модулей. Пусть $d \in \{-1, \ldots, n(p-1)\}$. Тогда ввиду равенства (12) RM_d отображается при изоморфизме (15) на аннулятор идеала Δ^{d+1} в FG_n. Известно, что этот аннулятор совпадает с $\Delta^{n(p-1)-d}$ (основная теорема из [13] или [11, теоремы 4.2.2, 3.2.2]). Этим объясняется то, что в определении и теореме 3.3.12 из [11] $\Delta^{n(p-1)-d}$ называется обобщённым кодом Рида — Маллера порядка d и длины p^n. Следовательно, если $d \in \{1, \ldots, n(p-1)\}$, то

$$\mathrm{HN}_p(n, d) = |\{\alpha \in (\Delta^{n(p-1)-d}/\Delta^{n(p-1)-d+1}) \setminus \{0\} : L(\alpha) = \{1\}\}| \tag{16}$$

ввиду замечания 2. Здесь

$$L(\xi + \Delta^{n(p-1)-d+1}) = L(\xi) = \{g \in G_n : (g-1)\xi \in \Delta^{n(p-1)-d+2}\}$$

для произвольного $\xi \in \Delta^{n(p-1)-d} \setminus \Delta^{n(p-1)-d+1}$; корректность определения очевидна.

Пусть $d \in \{1, \ldots, n(p-1)\}$ и H — подгруппа группы G_n. Положим

$$\gamma_d(H) = |\{\alpha \in (\Delta^{n(p-1)-d}/\Delta^{n(p-1)-d+1}) \setminus \{0\} : L(\alpha) = H\}|,$$
$$\Gamma_d(H) = |\{\alpha \in (\Delta^{n(p-1)-d}/\Delta^{n(p-1)-d+1}) \setminus \{0\} : L(\alpha) \supseteq H\}|.$$

Вычислим $\Gamma_d(H)$. Выберем базис $b = (b_1, \ldots, b_n)$ элементарной абелевой p-группы G_n так, чтобы $b_{m+1}, \ldots, b_n$ порождали подгруппу H ($m \in \{0, \ldots, n\}$). Пусть $\xi \in \Delta^{n(p-1)-d} \setminus \Delta^{n(p-1)-d+1}$ и $\alpha = \xi + \Delta^{n(p-1)-d+1}$. Так как $L(\alpha)$ является подгруппой группы G_n, включение $L(\alpha) \supseteq H$ выполняется тогда и только тогда, когда $(b_i - 1)\xi \in \Delta^{n(p-1)-d+2}$ для всех $i \in \{m+1, \ldots, n\}$. Непосредственно проверяется (с использованием равенства (14)), что последнее условие имеет место, если и только если $\xi \in U + \Delta^{n(p-1)-d+1}$, где

$$U = \langle\{\eta_j(b) : j = (j_1, \ldots, j_m, p-1, \ldots, p-1) \in J_n,\ j_1 + \ldots + j_m = m(p-1) - d\}\rangle.$$

Сумма U и $\Delta^{n(p-1)-d+1}$ прямая, поэтому отсюда следует, что $\Gamma_d(H) = |U| - 1$. Кроме того, $\dim U = \left\{ {m \atop m(p-1)-d} \right\}_p = \left\{ {m \atop d} \right\}_p$ (см. формулу (4)). Таким образом,

$$\Gamma_d(H) = p^{\left\{ {n - \operatorname{rank} H \atop d} \right\}_p} - 1. \tag{17}$$

4. Доказательства теорем 2 и 3

Через $\mathfrak{S}(G_n)$ будем обозначать частично упорядоченное по включению множество всех подгрупп группы G_n.

Доказательство теоремы 2. Пусть $n \geqslant 1$ и $d \in \{1, \ldots, n(p-1)\}$. Очевидно, что

$$\Gamma_d(X) = \sum_{Y \in \mathfrak{S}(G_n) \,|\, Y \supseteq X} \gamma_d(Y)$$

для любого $X \in \mathfrak{S}(G_n)$. Поэтому ввиду формулы (16) и формулы обращения Мёбиуса [6, утверждение 4.18, п. (ii)] получаем равенство

$$\mathrm{HN}_p(n, d) = \gamma_d(\{1\}) = \sum_{X \in \mathfrak{S}(G_n)} \mu(\{1\}, X)\Gamma_d(X), \tag{18}$$

где μ — функция Мёбиуса [6, подраздел B раздела 1 главы IV] частично упорядоченного множества $\mathfrak{S}(G_n)$. Известно, что

$$\mu(\{1\}, X) = (-1)^{\operatorname{rank} X} p^{\binom{\operatorname{rank} X}{2}} \tag{19}$$

для каждого $X \in \mathfrak{S}(G_n)$ [6, предложение 4.20, п. (iii)]. Следовательно,

$$\mathrm{HN}_p(n, d) = \sum_{k=0}^{n} (-1)^k p^{\binom{k}{2}} \left(p^{\left\{ {n-k \atop d} \right\}_p} - 1\right) \begin{bmatrix} n \\ k \end{bmatrix}_p \tag{20}$$

согласно формулам (17)–(19). Кроме того,

$$\sum_{k=0}^{n} (-1)^k p^{\binom{k}{2}} \begin{bmatrix} n \\ k \end{bmatrix}_p = \sum_{X \in \mathfrak{S}(G_n)} \mu(\{1\}, X) = 0 \tag{21}$$

ввиду (19) и [6, предложение 4.6] (напомним, что $n \geqslant 1$). Первое равенство в (6) непосредственно вытекает из равенств (20) и (21), второе следует из того, что

$$\begin{bmatrix} n \\ k \end{bmatrix}_p = \sum_{S \subseteq \{1,\ldots,n\} \,|\, |S|=k} p^{\sigma(S)-\binom{k+1}{2}}$$

для любого $k \in \{0,\ldots,n\}$, где $\sigma(S)$ — сумма всех элементов множества $S \subseteq \{1,\ldots,n\}$ [7, теорема 6.1]. При доказательстве второго равенства в (6) используется также очевидное свойство $\binom{k+1}{2} = \binom{k}{2} + k$. ■

Доказательство теоремы 3. Пусть $d \geqslant 1$ и $n \geqslant d/(p-1)$. Пусть также $\alpha \in (\Delta^{n(p-1)-d}/\Delta^{n(p-1)-d+1}) \setminus \{0\}$, причём $L(\alpha) \neq \{1\}$. Тогда $L(\alpha)$ содержит некоторую подгруппу ранга 1. Используя это замечание, а также формулы (4), (13), (16), (17) и (5), получаем

$$\begin{aligned} p^{\left\{{n \atop d}\right\}_p} - 1 - \mathrm{HN}_p(n,d) &= p^{\left\{{n \atop n(p-1)-d}\right\}_p} - 1 - \mathrm{HN}_p(n,d) = |(\Delta^{n(p-1)-d}/\Delta^{n(p-1)-d+1}) \setminus \{0\}| - \\ &- |\{\alpha \in (\Delta^{n(p-1)-d}/\Delta^{n(p-1)-d+1}) \setminus \{0\} \,|\, L(\alpha) = \{1\}\}| \leqslant \\ \leqslant \sum_{H \in \mathfrak{S}(G_n) \,|\, \operatorname{rank} H = 1} \Gamma_d(H) &= \begin{bmatrix} n \\ 1 \end{bmatrix}_p \left(p^{\left\{{n-1 \atop d}\right\}_p} - 1\right) = \frac{p^n - 1}{p-1}\left(p^{\left\{{n-1 \atop d}\right\}_p} - 1\right), \end{aligned}$$

откуда непосредственно следует неравенство (8).

Предположим теперь, что $d \geqslant 3$. Из (8) и (10) вытекает, что

$$1 - \frac{1}{p^{\left\{{n \atop d}\right\}_p}} - \frac{(p^n-1)(p^{\left\{{n-1 \atop d}\right\}_p} - 1)}{(p-1)p^{\left\{{n \atop d}\right\}_p}} \leqslant \frac{\mathrm{HN}_p(n,d)}{p^{\left\{{n \atop d}\right\}_p}} \leqslant 1. \tag{22}$$

Мы уже видели при доказательстве формулы (11), что $\left\{{n \atop d}\right\}_p \to +\infty$ при $n \to \infty$. Поэтому

$$\frac{1}{p^{\left\{{n \atop d}\right\}_p}} \to 0 \text{ при } n \to \infty. \tag{23}$$

Кроме того, непосредственно проверяется, что

$$\left\{{n \atop d}\right\}_p - \left\{{n-1 \atop d}\right\}_p = \sum_{i=1}^{p-1} \left\{{n-1 \atop d-i}\right\}_p \geqslant \left\{{n-1 \atop d-1}\right\}_p \geqslant \binom{n-1}{d-1}$$

(см. также [5, формула (1.13)]). Поэтому $\left\{{n \atop d}\right\}_p - \left\{{n-1 \atop d}\right\}_p - n \to +\infty$ при $n \to \infty$, так как $\binom{n-1}{d-1} - n \sim n^{d-1}/(d-1)!$ при $n \to \infty$ (напомним, что $d-1 \geqslant 2$). Следовательно,

$$\frac{(p^n-1)(p^{\left\{{n-1 \atop d}\right\}_p} - 1)}{p^{\left\{{n \atop d}\right\}_p}} \sim \frac{1}{p^{\left\{{n \atop d}\right\}_p - \left\{{n-1 \atop d}\right\}_p - n}} \to 0 \text{ при } n \to \infty. \tag{24}$$

Из (22)–(24) вытекает теперь, что $\mathrm{HN}_p(n,d)/p^{\left\{{n \atop d}\right\}_p} \to 1$ при $n \to \infty$. Таким образом, $\mathrm{HN}_p(n,d) \sim p^{\left\{{n \atop d}\right\}_p}$ при $n \to \infty$. ■

Автор благодарит анонимного рецензента, отзыв которого побудил автора к существенной переработке первоначальной версии настоящей работы.

ЛИТЕРАТУРА

1. *Черемушкин А. В.* Аддитивный подход к определению степени нелинейности дискретной функции // Прикладная дискретная математика. 2010. № 2(8). С. 22–33.
2. *Логачев О. А., Сальников А. А., Ященко В. В.* Невырожденная нормальная форма булевых функций // Доклады РАН. 2000. Т. 373. № 2. С. 164–167.
3. *Логачев О. А., Сальников А. А., Смышляев С. В., Ященко В. В.* Булевы функции в теории кодирования и криптологии. М.: ЛЕНАНД, 2015.
4. *MacWilliams J.* Orthogonal matrices over finite fields // Amer. Math. Monthly. 1969. V. 76. No. 2. P. 152–164.
5. *Бондаренко Б. А.* Обобщенные треугольники и пирамиды Паскаля, их фракталы, графы и приложения. Ташкент: Фан, 1990.
6. *Айгнер М.* Комбинаторная теория. М.: Мир, 1982.
7. *Кац В. Г., Чен П.* Квантовый анализ. М.: МЦНМО, 2005.
8. *Кузнецов Ю. В.* О числе невырожденных булевых форм // Труды лаборатории МГУ по математическим проблемам криптографии. 2000. С. 78–80.
9. *Анохин М. И.* О некоторых множествах групповых функций // Матем. заметки. 2003. Т. 74. № 1. С. 3–11.
10. *Jennings S. A.* The structure of the group ring of a p-group over a modular field // Trans. Amer. Math. Soc. 1941. V. 50. No. 1. P. 175–185.
11. *Циммерман К.-Х.* Методы теории модулярных представлений в алгебраической теории кодирования. М.: МЦНМО, 2011.
12. *Берман С. Д.* К теории групповых кодов // Кибернетика. 1967. № 1. С. 31–39.
13. *Hill E. T.* The annihilator of radical powers in the modular group ring of a p-group // Proc. Amer. Math. Soc. 1970. V. 25. No. 4. P. 811–815.

REFERENCES

1. *Cheremushkin A. V.* Additivnyy podkhod k opredeleniyu stepeni nelineynosti diskretnoy funktsii [An additive approach to defining the degree of nonlinearity of a discrete function]. Prikladnaya Diskretnaya Matematika, 2010, no. 2(8), pp. 22–33. (in Russian)
2. *Logachev O. A., Sal'nikov A. A., and Yashchenko V. V.* Nevyrozhdennaya normalnaya forma bulevykh funktsiy [The nondegenerate normal form of Boolean functions]. Doklady RAN, 2000, vol. 373, no. 2, pp. 164–167. (in Russian)
3. *Logachev O. A., Sal'nikov A. A., Smyshlyaev S. V., and Yashchenko V. V.* Bulevy funktsii v teorii kodirovaniya i kriptologii [Boolean Functions in Coding Theory and Cryptology]. Moscow, LENAND Publ., 2015. (in Russian)
4. *MacWilliams J.* Orthogonal matrices over finite fields. Amer. Math. Monthly, 1969, vol. 76, no. 2, pp. 152–164.
5. *Bondarenko B. A.* Generalized Pascal triangles and pyramids, their fractals, graphs, and applications. Santa Clara, CA, The Fibonacci Association, 1993.
6. *Aigner M.* Combinatorial Theory. Berlin et al., Springer Verlag, 1979.
7. *Kac V. and Cheung P.* Quantum Calculus. New York et al., Springer Verlag, 2002.
8. *Kuznetsov Yu. V.* O chisle nevyrozhdennykh bulevykh form [On the number of nondegenerate Boolean forms]. Trudy laboratorii MGU po matematicheskim problemam kriptografii, 2000, pp. 78–80. (in Russian)
9. *Anokhin M. I.* O nekotorykh mnozhestvakh gruppovykh funktsiy [On some sets of group functions]. Matem. Zametki, 2003, vol. 74, no. 1, pp. 3–11. (in Russian)
10. *Jennings S. A.* The structure of the group ring of a p-group over a modular field. Trans. Amer. Math. Soc., 1941, vol. 50, no. 1, pp. 175–185.

11. *Zimmermann K.-H.* Beiträge zur algebraischen Codierungstheorie mittels modularer Darstellungstheorie. Bayreuther mathematische Schriften, Heft 48, 1994. (in German)
12. *Berman S. D.* K teorii gruppovykh kodov [On the theory of group codes]. Kibernetika, 1967, no. 1, pp. 31–39. (in Russian)
13. *Hill E. T.* The annihilator of radical powers in the modular group ring of a p-group. Proc. Amer. Math. Soc., 1970, vol. 25, no. 4, pp. 811–815.

УДК 519.7

О ПОСТРОЕНИИ APN-ПЕРЕСТАНОВОК С ПОМОЩЬЮ ПОДФУНКЦИЙ[1]

В. А. Идрисова

Институт математики им. С. Л. Соболева СО РАН,
Новосибирский государственный университет, г. Новосибирск, Россия

Работа посвящена проблеме существования взаимно однозначных APN-функций от чётного числа переменных. Рассматриваются векторные 2-в-1 функции, изоморфные $(n-1)$-подфункциям APN-перестановок, которые могут быть построены с помощью специального алгоритма. Для того чтобы получить APN-перестановку, необходимо найти координатные булевы функции f, такие, что взаимно однозначная функция, полученная из данной $(n-1)$-подфункции и функции f, является APN-функцией. Вводится понятие ассоциированных перестановок и доказывается оценка на число таких координатных булевых функций для некоторой $(n-1)$-подфункции. Описан соответствующий алгоритм поиска взаимно однозначных APN-функций с помощью подфункций и координатных булевых функций.

Ключевые слова: *векторная функция, APN-функция, перестановка, подфункция.*

DOI 10.17223/20710410/41/2

ON CONSTRUCTING APN PERMUTATIONS USING SUBFUNCTIONS

V. A. Idrisova

Sobolev Institute of Mathematics, Novosibirsk State University, Novosibirsk, Russia

E-mail: vitkup@math.nsc.ru

Our subject for investigation is the problem of APN permutation existence for even number of variables. In this work, we consider 2-to-1 functions that are isomorphic to $(n-1)$-subfunctions of APN permutations. These 2-to-1 functions can be obtained with a special algorithm which searches for 2-to-1 APN functions that are potentially EA-equivalent to permutations. The algorithm is based on constructing special symbol sequences that are called admissible. It is known that $(n-1)$-subfunction of an APN permutation can be represented as a differentially 4-uniform 2-to-1 function that takes values from the half of the Boolean cube. Therefore, the following algorithm can be used to search for APN permutations. On the first step all the possible admissible sequences are constructed and we assign obtained sequences in order to find a differentially 4-uniform 2-to-1 function. Therefore, obtained function can be isomorphic to a $(n-1)$-subfunction of an APN permutation, so, this $(n-1)$-subfunction can be expanded to bijective APN function. In order to construct an APN permutation, we need to find all possible coordinate Boolean functions f such that the bijective function constructed from the given $(n-1)$-subfunction S and function f is APN.

[1]Работа поддержана грантами РФФИ № 18-31-00374 и 18-07-01394, программой фундаментальных научных исследований СО РАН № I.5.1., проект № 0314-2016-0017, Министерством образования и науки (задание № 1.12875.2018/12.1) и в рамках программы 5-100.

Unfortunately, the exhaustive search through the set of potential coordinate functions is computationally hard when $n \geqslant 7$, so, we need to estimate the number $n(S)$ of such coordinate Boolean functions. For a given bijective vectorial function F, we introduce an associated permutation $F^\star$ as follows. We split the set $\mathbb{F}_2^n$ into two disjoint subsets $\mathcal{F}_1$ and $\mathcal{F}_2$, fix integer k, indices $i_1, \ldots, i_k$, and index $j \notin \{i_1, \ldots, i_k\}$. Then the value $F^\star(x)$ is equal to $F(x)$ if $F(x) \in \mathcal{F}_1$ and $F^\star(x)$ is equal to $F(x) + e_j$ otherwise. We prove that $F^\star$ is an APN permutation if and only if F is an APN permutation. This fact allows us to obtain the necessary bound. We prove that if $n(S)$ is not equal to zero, then $n(S) \geqslant 2^n$.

Keywords: *vectorial function, APN function, permutation, subfunction.*

Введение

Стойкость современных симметричных шифров существенно зависит от характеристик, которыми обладают их компоненты, в частности S-блоки. В общем случае S-блок представляет собой векторную функцию из $\mathbb{F}_2^n$ в $\mathbb{F}_2^m$. Многие широко используемые блочные шифры, такие, например, как AES, ГОСТ Р 34.12-2015 (Кузнечик), Serpent, являются по своей структуре SP-сетями. Важным свойством, требуемым от S-блоков в SP-сетях, является их обратимость, то есть векторная функция, используемая в качестве S-блока, должна быть взаимно однозначной.

Появление метода дифференциального криптоанализа в 1990 г. потребовало от S-блоков новых характеристик. Так, было введено понятие функций с низкой дифференциальной равномерностью и APN-функций — векторных функций, обладающих оптимальной стойкостью к дифференциальному криптоанализу. Одна из самых важных задач в области APN-функций заключается в необходимости совместить свойства взаимной однозначности и минимально возможной дифференциальной равномерности в одном S-блоке. Для нечётных n существует множество конструкций таких функций, однако для чётных размерностей до сих пор известен лишь один пример взаимно однозначной APN-функции от шести переменных.

Данная работа посвящена проблеме поиска и построения взаимно однозначных APN-функций. Рассматриваются 2-в-1 векторные функции, обладающие низкой дифференциальной равномерностью, которые могут быть получены специальным алгоритмом. Показано, что с помощью данных 2-в-1 функций возможен поиск новых APN-перестановок. В п. 1 вводятся основные определения и аппарат APN-функций, а также рассматриваются некоторые известные результаты, связанные с проблемой существования APN-перестановок. В п. 2.1 описывается алгоритм поиска APN-перестановок с помощью подфункций и недостающих координатных булевых функций и формулируется основная задача работы. В п. 2.2 рассматривается дифференциально 4-равномерная 2-в-1 векторная функция и доказывается нижняя оценка числа координатных булевых функций, таких, что перестановка, полученная из исходной 2-в-1 векторной функции и данной координатной булевой функции, является APN-функцией. Данная оценка позволяет понять, как много таких координатных функций существует и, следовательно, насколько быстро найдётся APN-перестановка в процессе работы алгоритма. Для произвольной взаимно однозначной функции вводится понятие ассоциированной перестановки. Доказывается, что некоторая перестановка является APN-функцией тогда и только тогда, когда её ассоциированная перестановка также является APN-функцией.

1. Определения

1.1. Основные определения

Будем обозначать через $\mathbb{F}_2^n$ множество всех двоичных векторов длины n. Функция F из $\mathbb{F}_2^n$ в $\mathbb{F}_2^m$, где n и m — целые числа, называется *векторной булевой функцией*. Если $m = 1$, то функция F называется *булевой*. Произвольная векторная функция F может быть представлена как набор из m *координатных функций* $F = (f_1, \ldots, f_m)$, где f_i — булева функция от n переменных. Для произвольного ненулевого вектора $v \in \mathbb{F}_2^m$ линейная комбинация координатных функций $v \cdot F$ называется *компонентной функцией*. Любую векторную булеву функцию F можно единственным образом представить в виде *алгебраической нормальной формы* (АНФ):

$$F(x_1, \ldots, x_n) = \bigoplus_{k=1}^{n} \bigoplus_{1 \leqslant i_1 \leqslant \ldots \leqslant i_k \leqslant n} a_{i_1,\ldots,i_k} x_{i_1} \ldots x_{i_k} \oplus a_0,$$

где $a_{i_1,\ldots,i_k}, a_0 \in \mathbb{F}_2^m$. *Алгебраической степенью* функции F называется количество переменных в самом длинном слагаемом её АНФ, при котором коэффициент не равен нулю. Если алгебраическая степень F не превышает единицы, то F называется *аффинной*. Аффинная функция F называется *линейной*, если $F(\mathbf{0}) = \mathbf{0}$.

Векторная булева функция F называется *уравновешенной*, если она принимает каждое значение из $\mathbb{F}_2^m$ ровно 2^{n-m} раз, в частности, булева функция *уравновешена*, или *сбалансирована*, если она принимает каждое значение 2^{n-1} раз. В случае $n = m$ уравновешенная функция F называется *взаимно однозначной,* или *перестановкой*. *Производной* функции F по направлению a называется векторная функция $D_a F(x) = F(x+a)+F(x)$, где a — ненулевой вектор из $\mathbb{F}_2^n$. *Вектором значений* для векторной функции F называется вектор $\big(F(x^{(1)}), \ldots, F(x^{(2^n)})\big)$, где $x^{(1)}, \ldots, x^{(2^n)}$ — лексикографически упорядоченные двоичные векторы из $\mathbb{F}_2^n$. Векторная функция F из $\mathbb{F}_2^n$ в $\mathbb{F}_2^n$ называется *2-в-1 функцией*, если она принимает 2^{n-1} различных значений, каждое из которых встречается в векторе значений ровно два раза.

Мы можем сопоставить векторному пространству $\mathbb{F}_2^n$ конечное поле $\mathrm{GF}(2^n)$ и рассматривать векторную булеву функцию как функцию над этим полем. Тогда любая векторная функция F единственным образом представляется над $\mathrm{GF}(2^n)$ в следующей форме:

$$F(x) = \sum_{i=0}^{2^n-1} \lambda_i x^i, \ \lambda_i \in \mathrm{GF}(2^n).$$

Векторные булевы функции F и G называются *расширенно аффинно эквивалентными* (*EA-эквивалентными*), если $F = A_1 \circ G \circ A_2 + A$, где A_1, A_2 — взаимно однозначные аффинные функции над $\mathbb{F}_2^n$ и A — аффинная функция. Если функции F и G являются EA-эквивалентными и $A \equiv \mathbf{0}$, то F и G называются *аффинно эквивалентными*. Рассмотрим ещё одно отношение эквивалентности [1] на множестве векторных булевых функций. Две функции F и G называются *CCZ-эквивалентными*, если соответствующие множества $\{(x, y) \in \mathbb{F}_2^n \times \mathbb{F}_2^n : y = F(x)\}$ и $\{(x, y) \in \mathbb{F}_2^n \times \mathbb{F}_2^n : y = G(x)\}$ являются аффинно эквивалентными, т. е. если существует аффинный автоморфизм $A = (A_1, A_2)$, такой, что $y = F(x) \Leftrightarrow A_2(x, y) = G(A_1(x, y))$.

1.2. Взаимно однозначные APN-функции

Рассмотрим векторную функцию F из $\mathbb{F}_2^n$ в $\mathbb{F}_2^n$. Для векторов $a, b \in \mathbb{F}_2^n$, где $a \neq 0$, определим

$$\delta(a, b) = \big|\{x \in \mathbb{F}_2^n : F(x + a) + F(x) = b\}\big|, \quad \Delta_F = \max_{a \neq \mathbf{0}, b \in \mathbb{F}_2^n} \delta(a, b).$$

Функция F называется *дифференциально Δ_F-равномерной*. Чем меньше параметр Δ_F, тем выше стойкость шифра, содержащего F в качестве S-блока, к дифференциальному криптоанализу. Для векторных функций из $\mathbb{F}_2^n$ в $\mathbb{F}_2^n$ наименьшее значение Δ_F равно 2. В этом случае функция F называется *почти совершенно нелинейной* (*APN-функцией*). Данные понятия введены К. Ньюберг в работе [3]. Если F является APN-функцией, то любая EA-эквивалентная/CCZ-эквивалентная функция также является APN-функцией.

Наиболее известные представители класса APN-функций — это мономиальные функции, то есть функции вида $F(x) = x^d$ (таблица) над конечным полем $\mathrm{GF}(2^n)$. Известно [12], что APN-функции изучались ещё в СССР. Так, например, в 1964 г. В. А. Башевым и Б. А. Егоровым было доказано, что инверсия элемента поля является APN-функцией. Несмотря на то, что класс APN-функций активно изучается, в данной области по-прежнему большое количество открытых вопросов. Для дальнейшего изучения темы рекомендуем, например, обзоры [12 – 16], книги [17, 18] и т.д.

Известные мономиальные APN-функции вида x^d над полем $\mathbf{GF}(2^n)$

Название	Значение d	Условия	Ссылки
Голда	2^t+1	$(t,n)=1$	[2, 3]
Касами	$2^{2t}-2^t+1$	$(t,n)=1$	[4, 5]
Уолша	2^t+3	$n=2t+1$	[6, 7]
Нихо	$2^t+2^{t/2}-1$, t чётное $2^t+2^{(3t+1)/2}-1$, t нечётное	$n=2t+1$	[8, 9]
Инверсия	$2^{2t}-1$	$n=2t+1$	[3, 10]
Доббертина	$2^{4t}+2^{3t}+2^{2t}+2^t-1$	$n=5t$	[11]

Один из самых важных открытых вопросов в области APN-функций посвящён проблеме существования взаимно однозначных APN-функций, или *APN-перестановок*. В [19] выдвинута гипотеза (и доказана для случая $n=4$), что не существует APN-перестановок от чётного числа переменных. Однако в 2009 г. был найден первый пример взаимно однозначной APN-функции от шести переменных [20]. В работах [21, 22] рассматривается бесконечное семейство векторных функций, таких, что $\Delta_F \leqslant 4$, также содержащее APN-функцию Диллона, однако доказано, что это единственная APN-перестановка в данном семействе. До сих пор неизвестно, существуют ли другие APN-перестановки от шести переменных (неэквивалентные функции Диллона) и существуют ли они вообще для других чётных $n > 6$.

2. 2-в-1 функции как подфункции APN-перестановок

2.1. Алгоритм построения APN-перестановок с помощью $(n-1)$-подфункций

В работе [23] предложен новый способ построения 2-в-1 APN-функций, использующий символьные последовательности специального вида. Вектору значений 2-в-1 функции можно сопоставить символьную последовательность, такую, что одинаковым значениям соответствуют одни и те же символы, а различным значениям — разные символы. В том случае, когда F — APN-функция, такая последовательность называется *допустимой*. В [23] также описан алгоритм генерации всевозможных допустимых последовательностей.

Определение 1. Пусть F — векторная булева функция $F = (f_1, \ldots, f_n)$ из $\mathbb{F}_2^n$ в $\mathbb{F}_2^n$. Векторная функция F_j из $\mathbb{F}_2^n$ в $\mathbb{F}_2^{n-1}$ называется $(n-1)$-*подфункцией* функции F, если $F_j = (f_1, \ldots, f_{j-1}, f_{j+1}, \ldots, f_n)$ для некоторого индекса $j \in \{1, \ldots, n\}$.

Напомним, что векторному пространству $\mathbb{F}_2^n$ можно сопоставить целочисленное множество $\{0,\dots,2^n-1\}$, где каждое целое число является десятичным представлением двоичного числа, представляющего вектор. Тогда можно рассматривать $(n-1)$-подфункцию F_j из $\mathbb{F}_2^n$ в $\mathbb{F}_2^{n-1}$ как векторную функцию из $\mathbb{F}_2^n$ в $\mathbb{F}_2^n$, к которой в качестве первой координатной функции добавлена булева функция, тождественно равная нулю. Данная расширенная функция принимает значения только из множества $\{0,\dots,2^{n-1}-1\}$. Таким образом, множество $(n-1)$-подфункций изоморфно множеству таких векторных функций из $\mathbb{F}_2^n$ в $\mathbb{F}_2^n$. Далее будем рассматривать оба определения $(n-1)$-подфункции в зависимости от контекста.

Рассмотрим 2-в-1 функцию, которая принимает значения из $\{0,\dots,2^{n-1}-1\}$, обозначим множество таких 2-в-1 функций от n переменных через $\mathcal{T}_n$. Легко заметить, что любая $(n-1)$-подфункция взаимно однозначной векторной функции есть в точности функция из $\mathcal{T}_n$. В работе [23] доказаны следующие утверждения.

Теорема 1. Пусть F — взаимно однозначная APN-функция от n переменных. Тогда любая её $(n-1)$-подфункция является дифференциально 4-равномерной функцией из $\mathcal{T}_n$.

Вопрос характеризации APN-функции через её подфункции исследовался также в работе [24], где доказано, что любая подфункция APN-функции из $\mathbb{F}_2^{n-1}$ в $\mathbb{F}_2^{n-1}$ является APN-функцией или дифференциально 4-равномерной векторной функцией.

Напомним, что в данной работе дифференциально 4-равномерная функция — это функция, для которой значение $\max\limits_{a\neq\mathbf{0},b\in\mathbb{F}_2^n}\delta(a,b)$ равняется 4, то есть, согласно данному определению, APN-функции не являются дифференциально 4-равномерными. В некоторых работах встречается другое определение дифференциально δ-равномерных функций, которое включает функции меньших порядков дифференциальной равномерности, в частности APN-функции.

Теорема 2. Пусть F — взаимно однозначная APN-функция от n переменных. Тогда символьная последовательность, соответствующая вектору значений любой её $(n-1)$-подфункции, является допустимой последовательностью.

Из данных теорем следует, что любая APN-перестановка может быть получена из 2-в-1 дифференциально 4-равномерной функции, построенной при помощи допустимой последовательности. В [23] предложен следующий алгоритм для поиска взаимно однозначных APN-функций. На первом шаге строятся допустимые символьные последовательности и для каждой последовательности находится означивание, такое, что полученная 2-в-1 функция является дифференциально 4-равномерной. Следовательно, данная функция может быть изоморфна $(n-1)$-подфункции некоторой взаимно однозначной APN-функции и соответственно эта $(n-1)$-подфункция может быть достроена до APN-перестановки.

Без ограничения общности рассмотрим $(n-1)$-подфункцию $S=(s_1,\dots,s_{n-1})$, которая изоморфна 2-в-1 векторной функции из $\mathcal{T}_n$. Напомним, что любая такая функция является $(n-1)$-подфункцией некоторой взаимно однозначной векторной функции из $\mathbb{F}_2^n$ в $\mathbb{F}_2^n$. Для того чтобы её получить, нужно добавить к подфункции S недостающую координатную булеву функцию $f=s_n$ от n переменных, удовлетворяющую некоторым свойствам. Легко заметить, что функция f должна быть сбалансированной, так как искомая векторная функция взаимно однозначна. Поскольку $(n-1)$-подфункция $S=(s_1,\dots,s_{n-1})$ является 2-в-1 функцией, каждое из значений $0,1,\dots,2^{n-1}-1$ встретится ровно два раза. Соответственно на одну пару совпадающих значений подфункции S приходится либо упорядоченная пара значений $(0,1)$ недостающей булевой

функции, либо упорядоченная пара значений $(1, 0)$. Так как в векторе значений S имеется 2^{n-1} пар совпадающих значений, всего существует $2^{2^{n-1}}$ булевых функций $f = s_n$, таких, что $S = (s_1, \ldots, s_{n-1}, s_n)$ является взаимно однозначной функцией.

К сожалению, число $2^{2^{n-1}}$ уже при $n \geqslant 7$ очень велико для того, чтобы перебрать всевозможные варианты недостающих координатных булевых функций и проверить, является ли APN-функцией построенная взаимно однозначная функция $S = (s_1, \ldots, s_{n-1}, s_n)$. Поэтому, чтобы оценить, как быстро при переборе найдётся искомая взаимно однозначная APN-функция, необходимо найти количество тех булевых функций, которые дают именно APN-перестановку.

2.2. Оценка числа координатных булевых функций для $(n-1)$-подфункции

Для произвольного натурального n рассмотрим векторное пространство $\mathbb{F}_2^n$ и разобьём его на два равномощных непересекающихся подмножества $\mathbb{F}_2^n = V_1 \cup V_2$ следующим образом: пусть $V_1 = \{v \in \mathbb{F}_2^n : \mathrm{wt}(v) — \text{нечётное число}\}$ и $V_2 = \{v \in \mathbb{F}_2^n : \mathrm{wt}(v) — \text{чётное число}\}$, $\mathrm{wt}(v)$ — вес вектора v. Рассмотрим произвольную взаимно однозначную функцию $F = (f_1, \ldots, f_n)$. Зафиксируем k координатных функций $f_{i_1}, \ldots, f_{i_k}$ и разобьём $\mathbb{F}_2^n$ на два непересекающихся подмножества $\mathcal{F}_1$ и $\mathcal{F}_2$ следующим образом:

$$\mathcal{F}_j = \{(f_1(x), \ldots, f_n(x)) : f_{i_1}(x), \ldots, f_{i_k}(x) \in V_j,\ x \in \mathbb{F}_2^n\}, \quad j = 1, 2.$$

Пусть дано значение k, а также набор индексов $i_1, \ldots, i_k$ и индекс $j \notin \{i_1, \ldots, i_k\}$. Определим ассоциированную перестановку $F^\star$ следующим образом:

$$F^\star(x) = \begin{cases} F(x), & F(x) \in \mathcal{F}_1, \\ F(x) + \mathbf{e}_j, & F(x) \in \mathcal{F}_2. \end{cases}$$

Здесь $\mathbf{e}_j$ — вектор веса 1, содержащий 1 в j-й компоненте.

Теорема 3. Перестановка F является APN-функцией тогда и только тогда, когда перестановка $F^\star$ является APN-функцией.

Доказательство. Поскольку F — APN-функция, для любого ненулевого вектора $a \in \mathbb{F}_2^n$ её производная $D_aF(x) = F(x) + F(x + a)$ является 2-в-1 функцией. Зафиксируем произвольный вектор a' и рассмотрим вектор значений функции $D_{a'}F$. Он состоит из 2^{n-1} различных значений $B_{a'}F = \{b_1, \ldots, b_{2^{n-1}}\}$, каждое из которых встречается ровно два раза.

Рассмотрим перестановку $F^\star(x)$ и производную $D_{a'}F^\star$ для того же вектора a'. Без ограничения общности, для фиксированного аргумента x' возможны три случая:

1) $F(x') \in \mathcal{F}_1$ и $F(x' + a') \in \mathcal{F}_1$;
2) $F(x') \in \mathcal{F}_1$ и $F(x' + a') \in \mathcal{F}_2$;
3) $F(x') \in \mathcal{F}_2$ и $F(x' + a') \in \mathcal{F}_2$.

Рассмотрим первый случай. Поскольку $F(x')$ и $F(x' + a')$ лежат в $\mathcal{F}_1$, значение $D_{a'}F^\star(x') = D_{a'}F^\star(x' + a') = D_{a'}F(x') = D_{a'}F(x' + a')$ принадлежит $B_{a'}F$ и встречается два раза.

В третьем случае оба значения $F(x')$ и $F(x' + a')$ лежат в $\mathcal{F}_2$. Тогда значение производной $D_{a'}F^\star(x')$ равно значению производной $D_{a'}F(x')$, поскольку $D_{a'}F^\star(x') = F^\star(x') + F^\star(x' + a') = F(x') + \mathbf{e}_j + (x' + a') + \mathbf{e}_j = F(x') + (x' + a')$. Заметим, что $D_{a'}F^\star(x')$ совпадает со значением $D_{a'}F^\star(x' + a')$, следовательно, оно принадлежит $B_{a'}F$ и встречается два раза.

Чтобы доказать, что перестановка $F^\star$ является APN-функцией, необходимо показать, что значение производной, получаемое во втором случае, отлично от значений производной, получаемых в первом и третьем случаях, а также показать, что оно встретится в векторе значений функции $D_{a'}F^\star$ ровно два раза.

Докажем вторую часть необходимого условия. Поскольку $F(x')$ принадлежит $\mathcal{F}_1$, а $F(x'+a')$ принадлежит $\mathcal{F}_2$, значение производной $D_{a'}F^\star(x')$ равно значению $D_{a'}F(x')+$ $+\,\mathbf{e}_j$. Поскольку F — APN-функция, значение $D_{a'}F^\star(x')$ встречается ровно два раза, а значит, и $D_{a'}F^\star(x')$ встретится среди значений $D_{a'}F^\star(x')$ ровно два раза.

Заметим, что для любой пары $v_1, w_1 \in V_1$ и любой пары $v_2, w_2 \in V_2$ выполнено $v_i + w_i \in V_2$, $i = 1, 2$, а для любой пары $v_1 \in V_1$, $v_2 \in V_2$ выполнено $v_1 + v_2 \in V_1$. Следовательно, по построению множества $\mathcal{F}_1$ и $\mathcal{F}_2$ обладают аналогичными свойствами, а именно: для любой пары $v_1, w_1 \in \mathcal{F}_1$ и любой пары $v_2, w_2 \in \mathcal{F}_2$ выполнено $v_i + w_i \in \mathcal{F}_2$, $i = 1, 2$, а для любой пары $v_1 \in \mathcal{F}_1$, $v_2 \in \mathcal{F}_2$ выполнено $v_1 + v_2 \in \mathcal{F}_1$. Из этого следует, что значения производных в первом и третьем случаях принадлежат $\mathcal{F}_2$, а во втором случае — $\mathcal{F}_1$. Следовательно, поскольку $\mathcal{F}_1$ и $\mathcal{F}_2$ не пересекаются, производная $D_{a'}F^\star$ является 2-в-1 функцией. В силу произвольности выбора a' получаем, что производные функции $F^\star$ по всем направлениям являются 2-в-1 функциями, следовательно, $F^\star$ — APN-перестановка. ■

Пусть S является 2-в-1 дифференциально 4-равномерной функцией из $\mathbb{F}_2^n$ в $\mathbb{F}_2^n$, принимающей значения из множества $\{0, \ldots, 2^{n-1} - 1\}$, которая может быть представлена в виде $(n-1)$-подфункции $S = (s_1, \ldots, s_{n-1})$. Обозначим через $n(S)$ число таких булевых функций f от n переменных, что $H = (s_1, \ldots, s_{n-1}, f)$ является APN-перестановкой.

Теорема 4. Если значение $n(S)$ не равно нулю, то $n(S) \geqslant 2^n$.

Доказательство. Из теоремы 3 следует, что если $H = (s_1, \ldots, s_{n-1}, f)$ является APN-перестановкой для некоторой булевой функции f, то ассоциированная перестановка $H^\star$ для некоторого набора индексов $i_1, \ldots, i_k$ также является APN-функцией. Чтобы оценить число булевых функций f, таких, что $H = S \cup f$ является APN-перестановкой, необходимо найти количество ассоциированных перестановок $H^\star$, имеющих общую $(n-1)$-подфункцию $S = (s_1, \ldots, s_{n-1})$.

Чтобы определить ассоциированную перестановку $H^\star$, в общем случае необходимо задать значение k, набор индексов $i_1, \ldots, i_k$ и индекс $j \notin \{i_1, \ldots, i_k\}$. Заметим, что перестановки H и $H^\star$ имеют общую $(n-1)$-подфункцию $S = (s_1, \ldots, s_{n-1})$ тогда и только тогда, когда $j = n$. Докажем, что каждой ассоциированной перестановке соответствует своё число k и набор индексов $i_1, \ldots, i_k$; соответственно для построения перестановки необходимо и достаточно задать лишь эти параметры. Для этого требуется доказать, что ни для какого $k^* < k$ не найдётся непересекающихся множеств V_1^*, V_2^*, таких, что $\mathbb{F}_2^{k^*} = V_1^* \cup V_2^*$ и выполнено следующее свойство: для вектора длины k зафиксируем произвольные $t = k - k^*$ координат, тогда любой вектор $v^* \in V_i^*$ может быть получен выкалыванием этих t координат из некоторого вектора $v \in V_i$, $i = 1, 2$.

По построению все векторы из $\mathbb{F}_2^k$ нечётного веса лежат в V_1, а значит, все векторы стандартного базиса $\mathbf{e}_j$, $j = 1, \ldots, k$, также лежат в V_1. Заметим, что нулевой вектор лежит в V_2. Рассмотрим вектор $\mathbf{e}_j$, такой, что координата j встречается среди t фиксированных координат $i_1, \ldots, i_t$. После выкалывания координат $i_1, \ldots, i_t$ из $\mathbf{e}_j$ получается нулевой вектор, который принадлежит V_1^*, однако нулевой вектор также лежит и в V_2^*, поскольку он уже лежал в V_2 до операции выкалывания. Поскольку для

любого $k^* < k$ и любого $t = k - k^*$ такой вектор $\mathbf{e}_j$ найдётся, множества V_1^* и V_2^* всегда будут пересекаться для любых t и $k*$.

Следовательно, для каждого k множества $\mathcal{F}_1$ и $\mathcal{F}_2$, полученные из таких V_1 и V_2, не совпадут ни с какими множествами $\mathcal{F}_1'$ и $\mathcal{F}_2'$, определёнными для некоторого $k^* < k$. Это значит, что для каждой ассоциированной перестановки существует единственное число k и единственный набор индексов $i_1, \ldots, i_k$, и для того, чтобы найти число возможных ассоциированных перестановок для APN-перестановки H, нужно посчитать число возможных наборов координат $i_1, \ldots, i_k$ для каждого значения $k = 1, \ldots, n-1$. Их в точности $\sum\limits_{j=1}^{n-1} \binom{2^{n-1}}{j} = 2^{n-1} - 1$.

Напомним, что прибавление аффинной функции не меняет свойства функции быть APN, следовательно, если $H = (s_1, \ldots, s_{n-1}, f)$ является APN-перестановкой, то и $G = (s_1, \ldots, s_{n-1}, f + \mathbf{1})$ также ею является. Заметим, что прибавление единицы к последней координате эквивалентно тому, что мы меняем местами множества V_1 и V_2 при построении $\mathcal{F}_1$ и $\mathcal{F}_2$. Вместе с исходной функцией H имеем 2^{n-1} APN-перестановок, а поскольку к последней координате каждой перестановки ещё можем прибавить единицу, то всего получаем 2^n различных APN-перестановок, имеющих общую $(n-1)$-подфункцию $S = (s_1, \ldots, s_{n-1})$. Таким образом, если существует хотя бы одна булева функция f, такая, что $H = (s_1, \ldots, s_{n-1}, f)$ является APN-перестановкой и, следовательно, $n(S) \neq 0$, то $n(S) \geqslant 2^n$. ■

С помощью компьютерных вычислений установлено, что данная оценка является точной для $n = 3, 5$ и для всех рассмотренных спорадических примеров дифференциально 4-равномерных функций из $\mathcal{T}_n$ от шести переменных.

Остаются открытыми несколько интересных вопросов. Пусть S является дифференциально 4-равномерной функцией из $\mathcal{T}_n$. Может ли величина $n(S)$ в таком случае быть равной нулю? Другими словами, из любой ли 2-в-1 дифференциально 4-равномерной функции из $\mathbb{F}_2^n$ в $\mathbb{F}_2^n$, принимающей значения из множества $\{0, \ldots, 2^{n-1} - 1\}$, можно получить APN-перестановку? Для всех рассмотренных примеров $n(S)$ не равнялась нулю. Более того, с помощью компьютерных вычислений получено, что при $n = 4$ в $\mathcal{T}_n$ не существует дифференциально 4-равномерных функций. Напомним, что APN-перестановок при $n = 4$ также не существует.

Понятия EA-эквивалентности и CCZ-эквивалентности очень важны, когда мы говорим о поиске новых функций, поскольку найти новую APN-перестановку от шести переменных — это найти APN-перестановку, неэквивалентную APN-функции Диллона. Можно заметить, что утверждение теоремы 3 задаёт отношение эквивалентности на множестве APN-перестановок. Как эта эквивалентность соотносится с уже известными отношениями эквивалентности? Так, мы проверили несколько пар ассоциированных APN-перестановок от пяти и шести переменных, и все рассмотренные примеры являлись попарно CCZ-эквивалентными.

Заключение

В работе представлен новый способ поиска взаимно однозначных APN-функций. Описан алгоритм получения APN-перестановок из 2-в-1 дифференциально 4-равномерных векторных функций и координатных булевых функций. Доказана оценка числа таких координатных булевых функций для данной векторной функции. Данный результат позволяет оценить, насколько эффективен предложенный алгоритм поиска APN-перестановок. Введено понятие ассоциированной перестановки для взаимно од-

нозначной функции и доказано, что некоторая перестановка является APN-функцией тогда и только тогда, когда её ассоциированная перестановка также является APN-функцией. Сформулированы некоторые открытые вопросы, например про свойства отношения эквивалентности на множестве взаимно однозначных APN-функций, которое задаётся аппаратом ассоциированных перестановок.

Автор выражает благодарность Наталье Токаревой, Николаю Коломейцу и Анастасии Городиловой за обсуждения, ценные замечания и дополнения.

ЛИТЕРАТУРА

1. *Carlet C., Charpin P., and Zinoviev V.* Codes, bent functions and permutations suitable for DES-like cryptosystems // Des. Codes Cryptogr. 2000. V. 15. P. 125–156.
2. *Gold R.* Maximal recursive sequences with 3-valued recursive crosscorrelation functions // IEEE Trans. Inform. Theory. 1968. V. 14. P. 154–156.
3. *Nyberg K.* Differentially uniform mappings for cryptography // EUROCRYPT'93. LNCS. 1994. V. 765. P. 55–64.
4. *Kasami T.* The weight enumerators for several classes of subcodes of the second order binary Reed — Muller codes // Inform. Control. 1971. V. 18. P. 369–394.
5. *Janwa H. and Wilson R.* Hyperplane sections of Fermat varieties in P^3 in char. 2 and some applications to cyclic codes // Proc. AAECC-10. LNCS. 1993. V. 673. P. 180–194.
6. *Canteaut A., Charpin P., and Dobbertin H.* Binary m-sequences with three-valued crosscorrelation: a proof of Welch conjecture // IEEE Trans. Inform. Theory. 2000. V. 46. P. 4–8.
7. *Dobbertin H.* Almost perfect nonlinear functions over GF(2^n): the Welch case // IEEE Trans. Inform. Theory. 1999. V. 45. P. 1271–1275.
8. *Dobbertin H.* Almost perfect nonlinear functions over GF(2^n): the Niho case // Inform. Comput. 1999. V. 151. P. 57–72.
9. *Hollmann H., and Xiang Q.* A proof of the Welch and Niho conjectures on crosscorrelations of binary m-sequences // Finite Fields Appl. 2001. V. 7. P. 253–286.
10. *Beth T. and Ding C.* On almost perfect nonlinear permutations // EUROCRYPT'93. LNCS. 1993. V. 765. P. 65–76.
11. *Dobbertin H.* Almost perfect nonlinear power functions over GF(2^n): a new case for n divisible by 5 / eds. D. Jungnickel and H. Niederreiter. Finite Fields and Applications. Berlin; Heidelberg: Springer, 2001. P. 113–121.
12. *Глухов М. М.* О приближении дискретных функций линейными функциями // Математические вопросы криптографии. 2016. Т. 7. № 4. С. 29–50.
13. *Blondeau C. and Nyberg K.* Perfect nonlinear functions and cryptography // Finite Fields Appl. 2015. V. 32. P. 120–147.
14. *Carlet C.* Open questions on nonlinearity and on APN Functions // LNCS. 2015. V. 9061. P. 83–107.
15. *Pott A.* Almost perfect and planar functions // Des. Codes Cryptography. 2016. V. 78(1). P. 141–195.
16. *Тужилин М. Э.* Почти совершенные нелинейные функции // Прикладная дискретная математика. 2009. № 3. С. 14–20.
17. *Budaghyan L.* Construction and Analysis of Cryptographic Functions. Springer International Publishing, 2014. 168 p.
18. *Carlet C.* Vectorial Boolean functions for cryptography // Ch. 9 of the monograph "Boolean Methods and Models in Mathematics, Computer Science, and Engineering". Cambridge Univ. Press, 2010. P. 398–472.

19. *Hou X.-D.* Affinity of permutations of F_2^n // Discr. Appl. Math. Special Issue: Coding and Cryptography Archive. 2006. V. 154. P. 313–325.
20. *Browning K. A., Dillon J. F., McQuistan M. T., and Wolfe A. J.* An APN permutation in dimension six // 9-th Intern. Conf. Finite Fields and Their Applications Fq'09, Contemporary Math., AMS, 2010. V. 518. P. 33–42.
21. *Canteaut A., Duval S., and Perrin L.* A generalisation of Dillon's APN permutation with the best known differential and linear properties for all fields of size 2^{4k+2} // IEEE Trans. Inform. Theory. 2016. V. 63. P. 7575–7591.
22. *Perrin L., Udovenko A., and Biryukov A.* Cryptanalysis of a theorem: Decomposing the only known solution to the big APN problem // CRYPTO 2016. Part II. LNCS. 2016. V. 9815. P. 93–122.
23. *Idrisova V.* On an algorithm generating 2-to-1 APN functions and its applications to "the big APN problem" // Cryptography and Communications. 2018. P. 1–19.
24. *Городилова А. А.* Характеризация почти совершенно нелинейных функций через подфункции // Дискретная математика. 2015. № 27(3). С. 3–16.

REFERENCES

1. *Carlet C., Charpin P., and Zinoviev V.* Codes, bent functions and permutations suitable for DES-like cryptosystems. Des. Codes Cryptogr., 2000, vol. 15, pp. 125–156.
2. *Gold R.* Maximal recursive sequences with 3-valued recursive crosscorrelation functions. IEEE Trans. Inform. Theory, 1968, vol. 14, pp. 154–156.
3. *Nyberg K.* Differentially uniform mappings for cryptography. EUROCRYPT'93, LNCS, 1994, vol. 765, pp. 55–64.
4. *Kasami T.* The weight enumerators for several classes of subcodes of the second order binary Reed — Muller codes. Inform. Control., 1971, vol. 18, pp. 369–394.
5. *Janwa H. and Wilson R.* Hyperplane sections of Fermat varieties in P^3 in char. 2 and some applications to cyclic codes. Proc. AAECC-10, LNCS, 1993, vol. 673, pp. 180–194.
6. *Canteaut A., Charpin P., and Dobbertin H.* Binary m-sequences with three-valued crosscorrelation: a proof of Welch conjecture. IEEE Trans. Inform. Theory, 2000, vol. 46, pp. 4–8.
7. *Dobbertin H.* Almost perfect nonlinear functions over GF(2^n): the Welch case. IEEE Trans. Inform. Theory, 1999, vol. 45, pp. 1271–1275.
8. *Dobbertin H.* Almost perfect nonlinear functions over GF(2^n): the Niho case. Inform. Comput., 1999, vol. 151, pp. 57–72.
9. *Hollmann H. and Xiang Q.* A proof of the Welch and Niho conjectures on crosscorrelations of binary m-sequences. Finite Fields Appl., 2001, vol. 7, pp. 253–286.
10. *Beth T. and Ding C.* On almost perfect nonlinear permutations. EUROCRYPT'93, LNCS, 1993, vol. 765, pp. 65–76.
11. *Dobbertin H.* Almost perfect nonlinear power functions over GF(2^n): a new case for n divisible by 5. Eds. D. Jungnickel and H. Niederreiter. Finite Fields and Applications, Berlin, Heidelberg, Springer, 2001, pp. 113–121.
12. *Glukhov M. M.* O priblizhenii diskretnykh funktsiy lineynymi funktsiyami [On the approximation of discrete functions by linear functions]. Mat. Vopr. Kriptogr., 2016, vol. 7, iss. 4, pp. 29–50. (in Russian)
13. *Blondeau C. and Nyberg K.* Perfect nonlinear functions and cryptography. Finite Fields Appl., 2015, vol. 32, pp. 120–147.
14. *Carlet C.* Open questions on nonlinearity and on APN Functions. LNCS, 2015, vol. 9061, pp. 83–107.

15. *Pott A.* Almost perfect and planar functions. Des. Codes Cryptography, 2016, vol. 78(1), pp. 141–195.
16. *Tuzhilin M. E.* Pochti sovershennye nelineynye funktsii [APN-functions]. Prikladnaya Diskretnaya Matematika, 2009, no. 3, pp. 14–20. (in Russian)
17. *Budaghyan L.* Construction and Analysis of Cryptographic Functions. Springer International Publ., 2014. 168 p.
18. *Carlet C.* Vectorial Boolean functions for cryptography. Ch. 9 of the monograph "Boolean Methods and Models in Mathematics, Computer Science, and Engineering", Cambridge Univ. Press, 2010, pp. 398–472.
19. *Hou X.-D.* Affinity of permutations of F_2^n. Discr. Appl. Math. Special Issue: Coding and Cryptography Archive, 2006, vol. 154, pp. 313–325.
20. *Browning K. A., Dillon J. F., McQuistan M. T., and Wolfe A. J.* An APN permutation in dimension six. 9-th Intern. Conf. Finite Fields and Their Applications Fq'09, Contemporary Math., AMS, 2010, vol. 518, pp. 33–42.
21. *Canteaut A., Duval S., and Perrin L.* A generalisation of Dillon's APN permutation with the best known differential and linear properties for all fields of size 2^{4k+2}. IEEE Trans. Inform. Theory, 2016, vol. 63, pp. 7575–7591.
22. *Perrin L., Udovenko A., and Biryukov A.* Cryptanalysis of a theorem: Decomposing the only known solution to the big APN problem. CRYPTO 2016, Part II, LNCS, 2016, vol. 9815, pp. 93–122.
23. *Idrisova V.* On an algorithm generating 2-to-1 APN functions and its applications to "the big APN problem". Cryptography and Communications, 2018, pp. 1–19.
24. *Gorodilova A. A.* Characterization of almost perfect nonlinear functions in terms of subfunctions. Discr. Math. Appl., 2016, vol. 26(4), pp. 193–202.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ КРИПТОГРАФИИ

УДК 519.2

КРИТЕРИИ МАРКОВОСТИ АЛГОРИТМОВ БЛОЧНОГО ШИФРОВАНИЯ

О. В. Денисов

ООО «Инновационные телекоммуникационные технологии», г. Москва, Россия

Изучаются вероятностные модели блочных шифрсистем, в которых случайные раундовые ключи независимы и одинаково распределены. Они называются марковскими шифрами, если последовательность раундовых разностей образует простую однородную цепь Маркова. Уточнены и обобщены критерий и достаточное условие марковости моделей шифрсистем. Расширен класс марковских шифров, построенный в диссертации швейцарского учёного X. Lai. Получены достаточные условия, при которых формула для матриц вероятностей переходов разностей расширенного класса содержит тензорное произведение матриц вероятностей переходов S-блоков.

Ключевые слова: *марковские шифры, случайные подстановки, вероятности переходов разностей.*

DOI 10.17223/20710410/41/3

CRITERIA FOR MARKOV BLOCK CIPHERS

O. V. Denisov

Innovative Telecommunication Technologies, LLC, Moscow, Russia

E-mail: denisovOleg@yandex.ru

We study probabilistic models of block ciphers with random independent identically distributed round keys. We call they by Markov ciphers if sequence of round differentials is a simple homogeneous Markov chain. Criteria and sufficient condition for this property are adjusted and generalized. Particularly, we prove that, for an iterative r-round block cipher with group operation on the set $\mathcal{X}$ of blocks and round function g, the following four conditions are equivalent: 1) for any plaintext of two blocks (X, X^*), the sequence of random round differentials $\Delta X = X^* X^{-1}, \Delta X(1) = = X^*(1)X(1)^{-1}, \ldots, \Delta X(r) = X^*(r)X(r)^{-1}$ is a homogeneous Markov chain under any distribution of (X, X^*); 2) for all $a \in \mathcal{X} \setminus \{e\}$, the distribution of $g(ax)g(x)^{-1}$ doesn't depend on $x \in \mathcal{X}$; 3) $\forall a \in \mathcal{X}\setminus\{e\}, x \in \mathcal{X}$ $(g(ax)g(x)^{-1} \sim g(aX)g(X)^{-1})$ under any distribution of X; 4) $\forall x \in \mathcal{X}$ $(g(\Delta X\, x)g(x)^{-1} \sim g(\Delta X\, X)g(X)^{-1})$ under any distribution of $(X, \Delta X)$. The class of Markov ciphers constructed in Lai's dissertation is expanded. We give sufficient conditions under which formula for the transition probabilities matrix of the expanded class contains tensor product of S-box transition probabilities matrices.

Keywords: *Markov ciphers, random permutations, transition probabilities of differentials.*

Введение

Пусть на алфавите $\mathcal{X}$ входных и выходных блоков задана групповая операция $\odot$, $e \in \mathcal{X}$ — нейтральный элемент, X^{-1} — обратный к элементу X группы $(\mathcal{X}, \odot)$. Как принято в дифференциальном (разностном) анализе, величины вида $X_1 \odot X_2^{-1}$ будем называть *разностями* блоков X_1, X_2. Через $\mathcal{X}' = \mathcal{X} \setminus \{e\}$ обозначим множество «ненулевых» разностей, $M = |\mathcal{X}| - 1$ — его мощность. Через $S(\mathcal{X})$ обозначим множество всех подстановок на множестве $\mathcal{X}$.

Рассматривается итеративный r-раундовый алгоритм блочного шифрования (3), в котором раундовые ключи принимают значения из некоторого множества Γ, каждый набор раундовых ключей $\gamma \in \Gamma^r$ задает шифрующую подстановку $g_\gamma \in S(\mathcal{X})$. Шифруется двублочный текст (X, X^*).

Будем работать в вероятностной модели такого случайного выбора текста и набора раундовых ключей, что

$$\text{открытый текст } (X, X^*) \text{ и набор } \gamma = (\gamma_1, \dots, \gamma_r) \text{ независимы;} \tag{1}$$

$$\text{ключи } \gamma_1, \dots, \gamma_r \text{ независимы и одинаково распределены.} \tag{2}$$

Тогда, как известно из курса теории вероятностей, случайные последовательности промежуточных блоков

$$\begin{gathered} X = X(0),\ X(1) = g_{\gamma_1}(X(0)), \dots, X(r) = g_{\gamma_r}(X(r-1)) = Y, \\ X^* = X^*(0),\ X^*(1) = g_{\gamma_1}(X(0)), \dots, X^*(r) = g_{\gamma_r}(X^*(r-1)) = Y^*, \end{gathered} \tag{3}$$

последовательность их пар

$$(X, X^*),\ (X(1), X^*(1)), \dots, (X(r), X^*(r)) = (Y, Y^*) \tag{4}$$

и последовательность промежуточных подстановок

$$g_{(\gamma_1,\dots,\gamma_t)} = g_{\gamma_1} \dots g_{\gamma_t},\ 1 \leqslant t \leqslant r,$$

являются простыми (далее рассматриваем только такие) однородными цепями Маркова. Но не всякая функция от последовательности состояний цепи Маркова (и, в частности, разность компонент пар) образует цепь Маркова. Поэтому содержательно следующее определение.

Определение 1. При условиях (1) и (2) будем говорить, что итеративный алгоритм шифрования (3) является *марковским шифром* (относительно операции $\odot$ при заданных распределениях входной пары (X, X^*) и раундового ключа γ_1), если последовательность случайных *раундовых разностей*

$$\Delta X = X^* X^{-1},\ \Delta X(1) = X^*(1) X(1)^{-1},\ \dots,\ \Delta X(r) = X^*(r) X(r)^{-1} \tag{5}$$

образует однородную цепь Маркова.

В (5) (и далее там, где это удобно) знак операции $\odot$ опущен для краткости.

1. Критерии марковости последовательности раундовых разностей

Обозначим для краткости $g = g_{\gamma_1}$; распределение этой случайной подстановки совпадает с распределением остальных раундовых подстановок согласно (2).

В следующей теореме доказан критерий (7), близкий к условиям [1, 2], и новые критерии (8), (9), которые можно назвать *условиями инвариантности распределения разностей* подстановки g относительно фиксации входного блока и/или входной разности. Пункт 1 доказательства основан на идее [2] использования теоремы [3] об укрупнении состояний цепей Маркова.

Теорема 1. Пусть выполнены условия (1) и (2), $\mathbb{P} = \|p_{a,b}\|_{a,b\in\mathcal{X}'}$ — фиксированная стохастическая матрица. Тогда эквивалентны следующие четыре условия:

$$\begin{gathered}\text{последовательность (5) образует однородную цепь Маркова}\\ \text{с матрицей вероятностей переходов } \mathbb{P} \text{ при любом распределении } (X, X^*);\end{gathered} \tag{6}$$

$$\text{для всех } a \in \mathcal{X}' \text{ распределение } g(ax)g(x)^{-1} \text{ не зависит от выбора } x \in \mathcal{X}; \tag{7}$$

$$\forall a \in \mathcal{X}', x \in \mathcal{X}\ (g(ax)g(x)^{-1} \sim g(aX)g(X)^{-1}) \text{ при любом распределении } X; \tag{8}$$

$$\forall x \in \mathcal{X}\ (g(\Delta X\, x)g(x)^{-1} \sim g(\Delta X\, X)g(X)^{-1}) \text{ при любом распределении } (X, \Delta X). \tag{9}$$

При этом матрица $\mathbb{P}$ является дважды стохастической, $p_{a,b} = \mathsf{P}\{g(ax)g(x)^{-1} = b\}$, $a, b \in \mathcal{X}'$.

Доказательство.

1. Докажем эквивалентность условий (6) и (7). Условие (6) означает, что цепь Маркова (4) допускает такое укрупнение состояний, при котором все пары (x, x^*) с одинаковой разностью $a \in \mathcal{X}'$ объединяются в одно состояние $a' = \{(x, ax) : x \in \mathcal{X}\}$, состоящее из $|\mathcal{X}|$ пар. Заметим, что a' является смежным классом группы $(\mathcal{X} \times \mathcal{X}, \odot)$ (являющейся прямым произведением двух групп) по её «диагональной» подгруппе $\mathcal{X}_1 = \{(x, x) : x \in \mathcal{X}\}$, поскольку $a' = (e, a) \odot \mathcal{X}_1$.

По теореме 6.3.2 [3, с. 160] такое укрупнение возможно при любом начальном распределении исходной цепи и даёт цепь Маркова с матрицей переходных вероятностей $\mathbb{P}$ тогда и только тогда, когда при любых фиксированных $a, b \in \mathcal{X}'$ вероятности

$$p_{(x,x^*),b'} = \sum_{(y,y^*)\in b'} \mathsf{P}\{X(1) = y, X^*(1) = y^* \,|\, X = x, X^* = x^*\}$$

одинаковы для всех пар $(x, x^*) \in a'$ и равны $p_{a,b}$. Но для таких пар $x^* = ax^{-1}$ и можно преобразовать

$$p_{(x,x^*),b'} = \sum_{y\in\mathcal{X}} \mathsf{P}\{g(X) = y, g(\Delta X)g(X)^{-1} = b \,|\, X = x, \Delta X = a\} = \mathsf{P}\{g(ax)g(x)^{-1} = b\},$$

где в последнем переходе использовано условие независимости (1).

Итак, цепь Маркова (4) допускает укрупнение состояний при любом распределении (X, X^*) тогда и только тогда, когда правые части последнего равенства одинаковы и равны $p_{a,b}$ для всех $x \in \mathcal{X}$, т. е. выполнено условие (7).

2. Справедливо следующее утверждение: если случайная величина ξ является функцией от случайного вектора η и, возможно, некоторых других случайных величин, то условие инвариантности условных распределений $(\xi \,|\, \eta = c)$ относительно всех возможных фиксаций вектора η эквивалентно тому, что распределение ξ одинаково при любом распределении η. Действительно, необходимость условия инвариантности относительно фиксаций очевидна, а достаточность легко вытекает из формулы полной вероятности

$$\mathsf{P}\{\xi = b\} = \sum_c \mathsf{P}\{\eta = c\}\, \mathsf{P}\{\xi = b|\, \eta = c\} = \mathsf{P}\{\xi = b|\, \eta = c_0\}$$

для любого фиксированного c_0 из множества значений η.

Применяя это утверждение к случайной величине $\xi = g(aX)g(X)^{-1}$ при $\eta = X$, получаем равносильность (7) и (8). Применяя это утверждение к случайной величине $\xi = g(\Delta X\, X)g(X)^{-1}$ при $\eta = \Delta X$, получаем равносильность (7) и (9).

3. Полагая $x = e$ в (7), получаем формулу

$$p_{a,b} = \mathsf{P}\left\{g(a)g(e)^{-1} = b\right\}. \tag{10}$$

Из неё легко вытекает дважды стохастичность матрицы $\mathbb{P}$:

$$\sum_{a\in\mathcal{X}'} p_{a,b} = \sum_{a\in\mathcal{X}'} \mathsf{P}\{b\,g(e) = g(a)\} = \mathsf{P}\left\{b\,g(e) \in \mathcal{X}\setminus\{g(e)\}\right\},$$

что равно 1, поскольку $b\,g(e) \neq g(e)$. Теорема 1 доказана. ■

Рассмотрим класс шифров, у которых раундовый ключ может быть разделён на две независимые части $\gamma = (\delta, \varepsilon)$, первая из которых равномерно распределена на $\mathcal{X}$ и накладывается на шифруемый блок в соответствии с групповой операцией, а вторая определяет применяемую случайную подстановку f:

$$\gamma = (\delta,\varepsilon),\ g_\gamma(x) = f_\varepsilon(x\odot\delta), \quad \delta \sim U(\mathcal{X}) \text{ и } \varepsilon \text{ независимы}. \tag{11}$$

Схема шифрования пар блоков такими раундовыми подстановками иллюстрируется рис. 1.

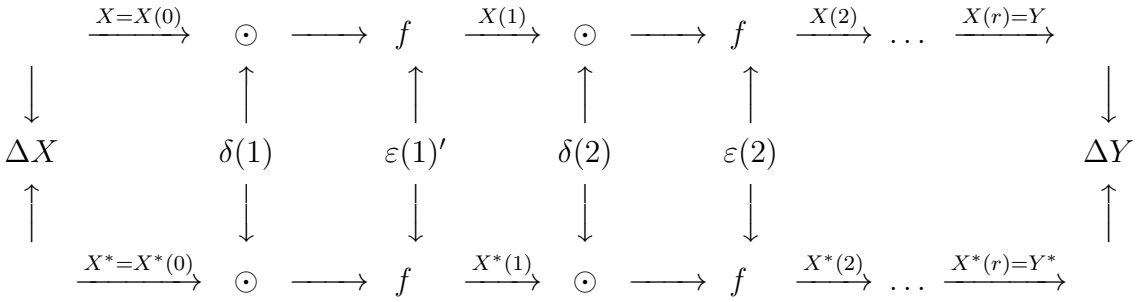

Рис. 1. Схема шифрования [4], обеспечивающая марковость алгоритма

Заметим, что схемы шифрования, использующие раундовые подстановки вида (11), близки к так называемым «key-alternating» шифрам, активно изучаемым в настоящее время.

Следствие 1. Пусть выполнены условия (1) и (2). Тогда условие (11) является достаточным для каждого из условий (6)–(9).

Доказательство. Согласно теореме 1, достаточно доказать, что условие (11) обеспечивает условие (7) инвариантности распределения разностей g_γ относительно фиксации входных блока и разности. Действительно, так как при любом $x \in \mathcal{X}$ случайная величина $y = x\delta$ также имеет равномерное распределение на $\mathcal{X}$ и не зависит от случайной подстановки f_ε, то распределение

$$g_\gamma(ax)g_\gamma(x)^{-1} = f_\varepsilon(ax\delta)f_\varepsilon(x\delta)^{-1} = f_\varepsilon(ay)f_\varepsilon(y)^{-1}$$

определяется распределением f_ε и значением a. ■

Заметим, что кроме случайных подстановок вида (11) условию (8) удовлетворяет равномерно распределённая подстановка $\sigma \sim U(S(\mathcal{X}))$, поскольку для неё распределение разности также не зависит от $x \in \mathcal{X}$ и даже от $a \in \mathcal{X}'$:

$$\mathsf{P}\{\sigma(ax)\sigma(x)^{-1}=b\}=\sum_{y\in\mathcal{X}}\mathsf{P}\{\sigma(ax)\sigma(x)^{-1}=b,\,\sigma(x)=y\}=$$
$$=\sum_{y\in\mathcal{X}}\mathsf{P}\{\sigma(x)=y,\sigma(ax)=by\}=(M+1)\frac{1}{(M+1)M}=\frac{1}{M}.$$

Обсудим возможную практическую пользу от формулы (10).

Начнем с *автономной вероятностной модели*: пусть $\mathscr{K}$ — множество всех ключей шифрования произвольной блочной шифрсистемы с некоторым заданным на нём вероятностным распределением. Это распределение при заданном алгоритме развёртывания ключа (АРК) индуцирует некоторое распределение на множестве Γ^r всех раундовых ключей, что с учётом распределения входной пары блоков позволяет найти матрицу $\mathbb{P}^{(r)}$ вероятностей переходов разностей (5) за r раундов. Для реализации разностной атаки различения с фиксированной входной разностью a требуется знание строки $\mathbb{P}_a^{(r)}$. Временная сложность метода вычисления $\mathbb{P}_a^{(r)}$ путём перебора ключей шифрования и вычисления переходов за r раундов всех пар блоков с фиксированной разностью a близка к сложности

$$|\mathscr{K}|(M+1)\odot 2r \tag{12}$$

операций вычисления значений раундовых подстановок.

Эта величина превосходит сложность тотального опробования, и в ряде работ по разностному анализу (в основном начиная с [5, 6]) рассматриваются *неавтономные вероятностные модели* шифрсистем, где АРК отсутствует, и раундовые ключи считаются независимыми одинаково распределёнными случайными величинами, т. е. выполняется условие (2). Марковские шифры удобны тем, что для них $\mathbb{P}^{(r)}=\mathbb{P}^r$ и сложность вычисления $\mathbb{P}^{(r)}$ близка к сложности вычисления $\mathbb{P}$, которая, согласно (13), меньше (12) примерно в $2r|\mathscr{K}|/|\Gamma|$ раз.

В следующей за [6] пионерской работе [1] перед теоремой 3 приведена («The (i,j) entry in Π is $\mathsf{P}\{\Delta Y(1)=\alpha_j\mid\Delta X=\alpha_i\}$») фактически формула (в наших обозначениях) $p_{a,b}=\mathsf{P}\{g(aX)g(X)^{-1}=b\}$ для элементов $\mathbb{P}$. Вычисление по ней фиксированной строки $\mathbb{P}_a$ может быть осуществлено перебором всех $x\in\mathcal{X}$, $\gamma\in\Gamma$ и добавлением произведений $\mathsf{P}\{X=x\}\,\mathsf{P}\{\gamma_1=\gamma\}$ к счётчику, соответствующему элементу $p_{a,b}$, $b=g_\gamma(x)^{-1}g_\gamma(xa)$. Для этого надо

$$(M+1)|\Gamma| \tag{13}$$

раз произвести две операции вычисления значения раундовой подстановки, две групповые операции и одну операцию вычисления обратного элемента в группе $(\mathcal{X},\odot)$. Осталось заметить, что для вычисления $\mathbb{P}_a$ по формуле (10) аналогичным способом достаточно $|\Gamma|$ операций вычисления значения раундовой подстановки и операций $\odot$, что примерно в $2M$ раз меньше сложности вычисления по формуле [1].

2. Шифрование параллельным набором S-боксов и подстановкой

Рассмотрим практически важный подкласс класса (11) марковских шифров, шифрование в перемешивающем слое которого осуществляется параллельным набором S-боксов, а затем действует подстановка π:

$$\begin{gathered}g_\gamma(x)=\pi\left(f_{1,\varepsilon_1}(x_1\odot_1\delta_1),\ldots,f_{m,\varepsilon_m}(x_m\odot_m\delta_m)\right),\\ \gamma=(\delta,\varepsilon),\ \delta=(\delta_1,\ldots,\delta_m),\ \varepsilon=(\varepsilon_1,\ldots,\varepsilon_m),\\ x=(x_1,\ldots,x_m)\in\mathcal{X}=\mathcal{X}_1\times\ldots\times\mathcal{X}_m,\end{gathered} \tag{14}$$

где i-й S-бокс действует на множестве $\mathcal{X}_i$, $1 \leqslant i \leqslant m$, и выбирается случайно (в общем случае) из некоторого фиксированного набора функций $(f_{i,j} : \mathcal{X}_i \to \mathcal{X}_i, j \in J_i)$, J_i — некоторое множество индексов. В этом случае $(\mathcal{X}, \odot)$ — прямое произведение групп $(\mathcal{X}_1, \odot_1), \ldots, (\mathcal{X}_m, \odot_m)$.

Здесь в следующей формуле для матрицы вероятностей переходов разностей возникает тензорное произведение матриц вероятностей переходов разностей меньшей размерности.

Теорема 2. Пусть при условиях (1), (2) в каждом раунде (14) компоненты $\delta_1, \ldots, \delta_m$, $\varepsilon_1, \ldots, \varepsilon_m$ раундового ключа γ независимы, $\delta_i \sim U(\mathcal{X}_i)$, $1 \leqslant i \leqslant m$, подстановка π является гомоморфизмом группы $(\mathcal{X}, \odot)$. Тогда при любом распределении (X, X^*) последовательность (5) образует однородную цепь Маркова с множеством состояний $\mathcal{X}$ и дважды стохастической матрицей

$$\mathbb{P} = (\mathbb{P}_1 \otimes \ldots \otimes \mathbb{P}_m)\Pi, \quad \mathbb{P}_i = \|p_{a,b}^{(i)}\|_{a,b \in \mathcal{X}_i},\ 1 \leqslant i \leqslant m,$$

где Π — матрица подстановки π; вероятности $p_{a,b}^{(i)} = \mathsf{P}\{f_{i,\varepsilon_i}(a \odot_i x) f_{i,\varepsilon_i}(x)^{-1} \odot_i = b\}$ вычислены в предположении независимости $x \sim U(\mathcal{X}_i)$ от ε_i.

Доказательство. Положим $f_\gamma(x) = \pi\left(f_{1,\varepsilon_1}(x_1), \ldots, f_{m,\varepsilon_m}(x_m)\right)$. Тогда $g_\gamma(x) = f_\gamma(x \odot \delta)$ и выполнено условие (11), поскольку, согласно условию теоремы, случайные векторы $\delta \sim U(\mathcal{X})$ и ε независимы.

Поэтому из следствия 1 и теоремы 1 получаем, что последовательность разностей образует однородную цепь Маркова с дважды стохастической матрицей переходных вероятностей и множеством состояний $\mathcal{X} \setminus \{e\}$, $e = (e_1, \ldots, e_m)$, e_i — нейтральные элементы групп $\mathcal{X}_i$. Нейтральный элемент e образует одноэлементный класс существенных состояний, и добавление его в множество состояний приводит к появлению соответствующего блока единичного размера на диагонали $\mathbb{P}$, не влияя на дважды стохастичность матрицы.

Найдём переходные вероятности. Для $a = (a_1, \ldots, a_m) \in \mathcal{X}$ и $b' = (b_1, \ldots, b_m) = \pi^{-1}(b) \in \mathcal{X}$ — прообраза блока b относительно действия π — имеем с учётом гомоморфности π

$$g_\gamma(ax) g_\gamma^{-1}(x) = \pi\left(f_\gamma(ax\delta)\right) \pi\left(f_\gamma(x\delta)\right)^{-1} = \pi\left(f_\gamma(ax\delta) f_\gamma(x\delta)^{-1}\right).$$

Отсюда с учётом независимости случайных векторов $(\delta_i, \varepsilon_i)$, $1 \leqslant i \leqslant m$, получаем

$$p_{a,b} = \mathsf{P}\left\{f_\gamma(ax\delta) f_\gamma(x\delta)^{-1} = b'\right\} =$$
$$= \prod_{1 \leqslant i \leqslant m} \mathsf{P}\left\{f_{i,\varepsilon_i}(a_i \odot_i x_i \odot_i \delta_i) f_{i,\varepsilon_i}(x_i \odot_i \delta_i)^{-1} = b_i\right\} = p_{a_1,b_1}^{(1)} \ldots p_{a_m,b_m}^{(m)}.$$

Теорема 2 доказана. ■

Рассмотрим случай, когда все компонентные группы одинаковы, т. е. $(\mathcal{X}, \odot)$ — прямое произведение m групп $(\mathcal{X}_1, \odot)$, S-боксы выбираются из набора $(f_j : \mathcal{X}_1 \to \mathcal{X}_1, j \in J)$:

$$g_\gamma(x) = \pi\left(f_{\varepsilon_1}(\delta_1 \odot x_1), \ldots, f_{\varepsilon_m}(\delta_m \odot x_m)\right), \quad x = (x_1, \ldots, x_m) \in \mathcal{X} = \mathcal{X}_1^m. \tag{15}$$

Если при этом индексы выбора одинаково распределены, то очевидно, что тензорное произведение матриц становится тензорной степенью одной матрицы.

Следствие 2. Пусть при условиях (1) и (2) в каждом раунде (15) раундовый ключ γ состоит из независимых в совокупности случайных величин $\delta_1, \ldots, \delta_m, \varepsilon_1, \ldots, \varepsilon_m$, где $\delta_i \sim U(\mathcal{X}_1)$, а ε_i одинаково распределены. Пусть также подстановка π с матрицей Π является гомоморфизмом группы $(\mathcal{X}, \odot)$. Тогда при любом распределении (X, X^*) последовательность (5) образует однородную цепь Маркова с дважды стохастической матрицей

$$\mathbb{P} = \mathbb{P}_1^{[m]}\Pi, \quad \mathbb{P}_1 = \| \mathsf{P}\{f_{\varepsilon_1}(ax) f_{\varepsilon_1}(x)^{-1} = b\|_{a,b\in\mathcal{X}_1},$$

где вероятности вычисляются в предположении независимости $x \sim U(\mathcal{X}_1)$ и ε_1.

Отметим, что если в (15) операция $\odot$ — сложение по модулю 2, то π является гомоморфизмом как в SP-сети, где π выбирается из группы перестановок координат векторов, так и в XSL-сети, где π выбирается из группы невырожденных линейных преобразований двоичных векторов.

Прокомментируем теорему 2. Известно [6, следствие 1], что при шифровании набором параллельных S-боксов и покоординатном аддитивном наложении раундового ключа вероятность перехода разности равна произведению вероятностей переходов её составных частей. Другими словами, матрица вероятностей перехода разностей при таком шифровании является тензорным произведением соответствующих матриц переходов частей разностей, соответствующих S-боксам [2; 7, с. 11]. Но при этом были сформулированы не все вероятностные условия (в частности, не было условия (1)), при которых такая матрица является матрицей вероятностей переходов разностной цепи Маркова; в теореме 2 этот недочёт устранён. С другой стороны, часть ключа, управляющая выбором индексов S-боксов, либо отсутствовала [7] (и тогда этот набор постоянен в раунде), либо считалась равномерно распределённой [4]. Теорема 2 показывает, что условие равномерной распределённости необязательно.

3. Обсуждение результатов

3.1. Об определениях марковского шифра

В [1] шифр называется марковским, если для всех $a, b \in \mathcal{X}'$

$$\begin{gathered}\text{при равномерном распределении } \gamma \\ \mathsf{P}\left\{g_\gamma(\Delta X\, X) g_\gamma(X)^{-1} = b \mid \Delta X = a,\, X = x\right\} \text{ не зависит от выбора } x \in \mathcal{X}.\end{gathered} \tag{16}$$

Заметим, что при условии (1) независимости входных блоков и раундовых ключей вторая строка условия (16) становится условием (7).

Доказано [1, теорема 2], что из (16) при условии (2) следует марковость последовательности разностей (5). Но в формулировке и в доказательстве теоремы имеются некоторые изъяны. В формулировке нет условия (1), без которого в общем случае даже последовательность пар блоков (4) не является цепью Маркова. В доказательстве показана марковость лишь отрезка из первых трёх разностей, не обоснована возможность исключения условия $\Delta X = \alpha$ при переходе от строки 3 к строке 4 в цепочке равенств. В тезисах [8] указан ещё один некорректный переход в доказательстве теоремы 2 [1].

В работе [2] выбран другой подход к доказательству марковости, основанный на теореме 6.3.2 [3] о возможности укрупнения состояний цепи Маркова, что позволило фактически доказать необходимость и достаточность условия (16) для марковости (5). В теореме 1 данной работы установлено, что доказательство [2] может быть проведено

в условиях, когда раундовые ключи могут выбираться не обязательно равновероятно. Но для получения критерия условие марковости последовательности разностей должно быть усилено требованием его выполнения при любом распределении входной пары блоков, иначе остаётся потенциальная возможность слабого укрупнения состояний [3, с. 169]. Доказанная дважды стохастичность $\mathbb{P}$ равносильна установленной в теореме 2 [1] стационарности равномерного распределения разностей. Заметим, что в нашей работе в качестве определения марковского шифра выбрано условие марковости (5) как более естественное по сравнению с близким к нему условием (16).

Достаточное условие (11) нашей работы является ослаблением условия [4, теорема 6], поскольку не требует равномерности распределения второй части раундового ключа. При этом доказательство его достаточности значительно проще доказательства теоремы 6 [4, с. 44]. До работы [4] марковость моделей шифрсистем доказывалась лишь в частных случаях для DES [6, лемма 1, теорема 1], PES [1, лемма 1], а также для шифрсистемы IPES, позже названной IDEA [4, с. 57]. Заметим, что условие (11) позволяет рассматривать другие вероятностные модели IDEA, в которых часть $(Z_5^{(1)}, Z_6^{(1)})$ [4, с. 65] раундового ключа не обязательно равномерно распределена, что может способствовать построению атак на слабые части ключа.

3.2. О критике марковской модели

Марковская модель блочных шифрсистем активно используется в англоязычной литературе с начала 1990-х гг. Примерно с 2006 г. интерес к ней стали проявлять украинские авторы [9 – 11], а в последние годы и российские. Вместе с тем имеются и критики этой модели, поскольку она, естественно, далека от совершенства. Они обычно выдвигают следующий аргумент:

«в реальности ключ всегда один и фиксирован, а не случаен». (17)

Приведём контраргументы.

1. Существует ряд ситуаций, в которых наблюдаются несколько шифртекстов, выработанных на разных ключах шифрования, например:

а) ключи получены из одного долговременного и различных разовых ключей;
б) ключи получены из фиксированного ключа и различных известных инициализирующих векторов (tweakable ciphers);
в) связанные ключи.

2. При некоторых алгоритмах развертывания ключа раундовые ключи, полученные на первых раундах, слабо зависимы, т. е. они близки к случайным независимым равновероятным векторам при случайном равновероятном выборе ключа шифрования. Модель случайного выбора ключа позволяет привлекать аппарат теории вероятностей для изучения характеристик «типичной» случайной подстановки, реализуемой блочной шифрсистемой, а также для изучения распределения вероятностей отклонений от таких характеристик.

3. Очевидно, аргумент (17) относится не к марковским шифрам как таковым, а вообще к неавтономным вероятностным моделям блочных шифрсистем. Но, например, в теории поточных шифрсистем давно используется и не вызывает возражений аналогичная модель для фильтрующего генератора, когда линейная рекуррентная последовательность, определяемая ключевым начальным заполнением, заменяется на последовательность независимых равновероятных битов. Для блочных шифрсистем увеличивается только размерность задачи и последовательность раундовых ключей,

определяемая ключом шифрования, заменяется на последовательность независимых равновероятных двоичных векторов.

Теория марковских шифров, как уже сказано выше, всего лишь изучает условия, при которых пары шифруемых блоков могут быть укрупнены в их разности с сохранением марковского свойства, которое даёт затем удобное равенство Колмогорова — Чепмена $\mathbb{P}^{(r)} = \mathbb{P}^r$ для матрицы переходных вероятностей за r тактов.

Подводя итог, перечислим достоинства (+) неавтономных моделей и некоторые их недостатки (−) как естественные продолжения достоинств:

(+) возможность исследования вероятностей переходов блоков и пар блоков за r раундов методами цепей Маркова, простое вычисление $\mathbb{P}^{(r)} = \mathbb{P}^r$; (−) возникновение вопроса об адекватности модели;

(+) результаты анализа устойчивы к модификации АРК; (−) теряется информация о возможной связи между раундовыми ключами;

(+) в марковских алгоритмах шифрования: возможность сокращения в два раза размерности состояния при переходе в неавтономной модели от пар блоков к их разностям.

Автор выражает признательность Б. А. Погорелову за постановку задачи и внимание к работе и Ф. М. Малышеву, чьи замечания способствовали существенному упрощению формулировок и доказательств результатов.

ЛИТЕРАТУРА

1. *Lai X., Massey J., and Murphy S.* Markov ciphers and differential cryptanalysis // Eurocrypt-1991. LNCS. 1991. V. 547. P. 17–38.
2. *Погорелов Б. А., Пудовкина М. А.* Разбиения на биграммах и марковость алгоритмов блочного шифрования // Математические вопросы криптографии. 2017. Т. 8. Вып. 1. С. 107–142.
3. *Кемени Дж., Снелл Дж.* Конечные цепи Маркова. М.: Наука, 1970. 271 с.
4. *Lai X.* On the Design and Security of Block Ciphers: dissertation for the degree of Doctor of Technical Sciences. Swiss Federal Institute of Technology, Zurich, 1992. 118 p.
5. *Biham E. and Shamir A.* Differential cryptanalysis of DES-like cryptosystems // CRYPTO-1990. LNCS. 1991. V. 537. P. 2–21.
6. *Biham E. and Shamir A.* Differential cryptanalysis of DES-like cryptosystems // J. Cryptology. 1991. V. 4. No. 1. P. 3–72.
7. *Глухов М. М.* О рассеивающих линейных преобразованиях для блочных шифрсистем // Математические вопросы криптографии. 2011. Т. 2. Вып. 2. С. 5–40.
8. *Дрелихов В. О., Никифоров М. С.* О марковских свойствах усредненных разностных характеристик итеративных блочных шифров // Тез. докл. на конф. РУСКРИПТО, 2017. `www.ruscrypto.ru/accociation/archive/rc2017.html`
9. *Алексейчук А. Н.* Верхние границы параметров, характеризующих стойкость немарковских блочных шифров относительно методов разностного и линейного криптоанализа // Науково-технічний журнал «Захист інформації». 2006. Вып. 3. С. 20–28.
10. *Ковальчук Л. В.* Обобщенные марковские шифры: построение оценок практической стойкости к дифференциальным атакам // Сб. материалов 2-й Междунар. научн. конф. по проблемам безопасности и противодействия терроризму, 25–26 октября 2006 г. М.: МЦНМО, 2006.
11. *Лисицкая И. В., Долгов В. И.* Блочные симметричные шифры и марковские процессы // Прикладная радиоэлектроника. 2012. Т. 11. Вып. 2. С. 137–143.

REFERENCES

1. *Lai X., Massey J., and Murphy S.* Markov ciphers and differential cryptanalysis. Eurocrypt-1991, LNCS, 1991, vol. 547, pp. 17–38.
2. *Pogorelov B. A. and Pudovkina M. A.* Razbieniya na bigrammah i markovost' algoritmov blochnogo shifrovaniya [Partitions on bigrams and Markov property of block ciphers]. Matematicheskie Voprosy Kriptografii, 2017, vol. 8, iss. 1, pp. 107–142. (in Russian)
3. *Kemeny J. G. and Snell J. L.* Finite Markov Chains. The University Series in Undergraduate Mathematics. Princeton, Van Nostrand, 1960. 271 p.
4. *Lai X.* On the Design and Security of Block Ciphers: dissertation for the degree of Doctor of Technical Sciences. Swiss Federal Institute of Technology, Zurich, 1992. 118 p.
5. *Biham E. and Shamir A.* Differential cryptanalysis of DES-like cryptosystems. CRYPTO-1990, LNCS, 1991, vol. 537, pp. 2–21.
6. *Biham E. and Shamir A.* Differential cryptanalysis of DES-like cryptosystems. J. Cryptology, 1991, vol. 4, no. 1, pp. 3–72.
7. *Gluhov M. M.* O rasseivayushchih linejnyh preobrazovaniyah dlya blochnyh shifrsistem [On mixing linear transforms for block ciphers]. Matematicheskie Voprosy Kriptografii, 2011, vol. 2, iss. 2, pp. 5–40. (in Russian)
8. *Drelikhov V. O. and Nikiforov M. S.* O markovskih svojstvah usrednennyh raznostnyh harakteristik iterativnyh blochnyh shifrov [On Markov properties of averaged difference characteristics of iterative block ciphers]. Proc. RUSCRIPTO-2017, `www.ruscrypto.ru/accociation/archive/rc2017.html` (in Russian)
9. *Aleksejchuk A. N.* Verhnie granicy parametrov, harakterizuyushchih stojkost' nemarkovskih blochnyh shifrov otnositel'no metodov raznostnogo i linejnogo kriptoanaliza [Upper limits of parameters characterizing the stability of non-Markov block ciphers with respect to the methods of difference and linear cryptanalysis]. Naukovo-tekhnichnij zhurnal «Zahist Informacii», 2006, iss. 3, pp. 20–28. (in Russian)
10. *Koval'chuk L. V.* Obobshchennye markovskie shifry: postroenie ocenok prakticheskoj stojkosti k differencial'nym atakam [Generalized Markov ciphers: construction of estimates of practical resistance to differential attacks]. Proc. 2nd Intern. Conf. on Security and Counter-Terrorism, October 25–26, 2006. Moscow, MCCME Publ., 2006. (in Russian)
11. *Lisickaya I. V. and Dolgov V. I.* Blochnye simmetrichnye shifry i markovskie processy [Block symmetric ciphers and Markov processes]. Prikladnaya Radioehlektronika, 2012, vol. 11, iss. 2, pp. 137–143. (in Russian)

УДК 003.26; 004.056.55; 512.54

МЕТОД НЕЛИНЕЙНОГО РАЗЛОЖЕНИЯ ДЛЯ АНАЛИЗА КРИПТОГРАФИЧЕСКИХ СХЕМ, ИСПОЛЬЗУЮЩИХ АВТОМОРФИЗМЫ ГРУПП[1]

В. А. Романьков, А. А. Обзор

Омский государственный университет им. Ф. М. Достоевского, г. Омск, Россия

Показано применение метода нелинейного разложения для криптографического анализа на примере двух схем, которые используют автоморфизмы группы. При некоторых ограничениях на группу, выбранную в качестве платформы шифрования, данный метод работает эффективно и позволяет раскрывать секретную информацию (распределяемый ключ или пересылаемое сообщение) без решения алгоритмически сложных проблем, заложенных в основу схемы. Такими являются нетеровы группы, для которых эффективно решается проблема поиска вхождения элемента в заданную подгруппу. В частности, этим свойством обладают конечно порожденные нильпотентные или, более общо, полициклические группы, часто рекомендуемые в качестве платформ в современной алгебраической криптографии.

Ключевые слова: *криптография, криптоанализ, распределение ключа, нелинейное разложение, проблема поиска вхождения.*

DOI 10.17223/20710410/41/4

A NONLINEAR DECOMPOSITION METHOD IN ANALYSIS OF SOME ENCRYPTION SCHEMES USING GROUP AUTOMORPHISMS

V. A. Roman'kov, A. A. Obzor

Dostoevskii Omsk State University, Omsk, Russia

E-mail: romankov48@mail.ru, obzor2503@gmail.com

This paper shows how the nonlinear decomposition method, that had been invented by the first author, works against two cryptographic schemes based on group automorphisms. In some cases we can find the secret data and break the scheme without solving the algorithmic problem on which scheme is based. More exactly, let G be a group and A be a finitely generated subgroup of the automorphism group $\mathrm{Aut}(G)$. Suppose, that the membership search problem for G is efficiently solvable for any subgroup of the form $\langle g^A\rangle$ generated by the all images of g under automorphisms of A, and every subgroup $\langle g^A\rangle$ is finitely generated. Then there exists an efficient algorithm to construct a finite generating set of $\langle g^A\rangle$ and the nonlinear decomposition method can be applied. In particular, if the elements $g, f = g^\alpha, h = f^\beta \in G$ are public, $\alpha, \beta \in \mathrm{Aut}(G)$, $\alpha\beta = \beta\alpha$, and α, β are private, then one can efficiently compute h^α without computing α or β. The method efficiently works for a Noetherian group with efficiently solvable membership search problem. In particular, finitely generated nilpotent (more generally, polycyclic) groups, that are frequently used in the modern algebraic cryptography, share this property.

[1]Исследование выполнено за счёт гранта Российского научного фонда (проект № 16-11-10002).

Keywords: *cryptography, cryptanalysis, key exchange, nonlinear decomposition, membership search problem.*

Введение

В данной работе рассматривается применение метода нелинейного разложения, предложенного первым автором в [1], на примере двух схем: схемы передачи ключа Махалонобиса [2], являющейся алгебраическим аналогом классической схемы Масси — Омуры, и алгебраического аналога классического протокола Эль-Гамаля, использующего автоморфизмы [3, 4] (относительно классических версий см., например, [5]). В отличие от метода линейного разложения, теоретические основы которого изложены в [6] (см. также [7, 8]), метод нелинейного разложения не предполагает наличия структуры векторного пространства у платформы, на которой строится схема.

В [2] А. Махалонобис предложил два протокола распределения ключа, один из которых является алгебраической версией протокола Диффи — Хеллмана, другой — алгебраической версией протокола Масси — Омуры. В качестве платформы шифрования предлагались конечно порождённые нильпотентные группы и их автоморфизмы специального вида. Оба протокола проанализированы в [6] в предположении, что используемая группа линейна. Методом линейного разложения установлено, что в этом случае протоколы уязвимы. Конечно порождённые нильпотентные группы всегда допускают точное представление матрицами, но размерность матриц может оказаться слишком большой для проведения такого анализа. Поэтому возникла потребность поиска других методов.

В [1] первым автором дан криптографический анализ первого из этих протоколов (аналога протокола Диффи — Хеллмана), показана его уязвимость. При этом не была использована детализация групп и автоморфизмов из [2]. Применён новый метод нелинейного анализа, основанный на эффективной разрешимости в группе проблемы вхождения элемента в подгруппу. Отмечено, что такая проблема эффективно разрешима в полициклических, а значит, и в конечно порождённых нильпотентных группах. В последнем классе большинство основных алгоритмических проблем разрешаются алгоритмами с низкой сложностью. Об этом говорится далее более подробно. В том числе существует алгоритм низкой сложности, решающий проблему поиска вхождения элемента в подгруппу.

В настоящей работе представлен криптографический анализ второго из протоколов работы [2] — алгебраической версии протокола Масси — Омуры. В предположениях автора протокола он оказывается уязвимым относительно анализа методом нелинейного разложения. Аналогичный анализ проведён также для алгебраической версии протокола Эль-Гамаля.

Метод нелинейного разложения используется также для криптографического анализа алгебраической версии протокола Эль-Гамаля и алгебраической версии системы MOR, предложенной в [3, 4]. Доказано, что эти версии также уязвимы.

Через $\langle g_1, \ldots, g_n \rangle$ будем обозначать подгруппу рассматриваемой группы, порождённую элементами $g_1, \ldots, g_n$. Выражение $H \leqslant G$ означает, что H — подгруппа группы G.

Напомним, что проблема поиска, соответствующая классической проблеме вхождения (поиска вхождения элемента в подгруппу), ставится следующим образом: для данной группы G, её подгруппы $H = \langle h_1, \ldots, h_k \rangle$ и заданного элемента $g \in H$ найти слово $u(x_1, \ldots, x_k)$, такое, что $g = u(h_1, \ldots, h_k)$. Если подгруппа H фиксирована, рассматриваем соответствующую проблему поиска вхождения элемента $g \in H$ в H.

Здесь и далее предполагается, что используется язык комбинаторной теории групп, в котором группы задаются своими представлениями через порождающие элементы и определяющие соотношения, элементы записываются словами от порождающих элементов. Допустимы также матричные задания групп и их подгрупп над полями, в которых основные операции эффективно выполнимы. Почти во всех известных случаях группы, предлагаемые как платформы для криптографических схем, задаются именно таким способом. Как правило, требуется наличие в группах нормальных форм записи элементов, операции над которыми также должны быть эффективными.

Пусть G — алгебраическая система (группа, полугруппа, кольцо и т. п.), $\mathrm{Aut}(G)$ — её группа автоморфизмов. Через g^α будем обозначать образ элемента $g \in G$ относительно автоморфизма $\alpha \in \mathrm{Aut}(G)$. Другими словами, $g^\alpha = \alpha(g)$. В данной работе мы ограничиваемся рассмотрением групп, но предлагаемый метод криптографического анализа при соответствующих условиях может быть применён и к произвольным алгебраическим системам.

Если A — подгруппа группы автоморфизмов $\mathrm{Aut}(G)$ группы G и $v \in G$, то $v^A = \{v^\alpha | \alpha \in A\}$ обозначает орбиту относительно A, порождённую элементом v.

1. Лемма о построении порождающего множества

Пусть G — группа, $A = \langle \alpha_1, \ldots, \alpha_n \rangle$ — подгруппа группы автоморфизмов $\mathrm{Aut}(G)$.

Лемма 1. Если в группе G эффективно разрешима проблема поиска вхождения элемента в подгруппу $\langle g^A \rangle$ и подгруппа $\langle g^A \rangle$ конечно порождена, то существует алгоритм построения её конечного порождающего множества.

Доказательство. Считаем для простоты, что множество $\{\alpha_1, \ldots, \alpha_n\}$ замкнуто относительно взятия обратных элементов. Упорядочим все конечные наборы вида $\alpha_{i_1} \ldots \alpha_{i_t}$, $t = 0, 1, \ldots$, в соответствии с длиной и лексикографическим порядком. Полученная последовательность содержит записи всех элементов множества g^A. Элемент 1 соответствует набору длины 0. Через L_i обозначим часть полученной последовательности, соответствующую наборам длины i; $M_i = g^{L_i}$ — индуцированно упорядоченный набор элементов вида g^α, $\alpha \in L_i$. Последовательность множеств M_i, $i = 0, 1, \ldots$, очевидно, содержит все элементы множества g^A, а значит, и какое-то конечное порождающее множество подгруппы $\langle g^A \rangle$.

Опишем процесс построения конечного порождающего множества. Положим $K_0 = M_0 = \{g\}$. Добавляем к K_0 по очереди элементы $g^{\alpha_j} \in M_1$, для которых $g^{\alpha_j} \notin \langle g, g^{\alpha_1}, \ldots, g^{\alpha_{j-1}} \rangle$. Другими словами, добавляем элементы, не принадлежащие подгруппе, порождённой ранее выбранными элементами. Шаг заканчивается, когда будут просмотрены все элементы из M_1. В результате получим эффективно построенное подмножество $K_1 = \{g, g^{\alpha_{l_1}}, \ldots, g^{\alpha_{l_s}} : l_1 < \ldots < l_s\}$ множества $K_0 \cup M_1$. При этом $\langle g^A \rangle = \langle K_1, \bigcup_{i=2}^{\infty} M_i \rangle$.

Если $K_0 = K_1$, то алгоритм завершает свою работу, констатируя, что $\langle g^A \rangle = \langle K_0 \rangle = \langle g \rangle$ — циклическая группа. Действительно, в этом случае подгруппа $\langle g \rangle$ очевидно инвариантна относительно любого автоморфизма α_i, $i = 1, \ldots, n$, значит, и любого автоморфизма из A.

Если $K_0 \neq K_1$, продолжаем описанный процесс, добавляя к K_1 поочередно элементы из M_2, не входящие в подгруппу, порождённую уже выбранными элементами. При этом можно сразу убрать из рассмотрения элементы вида $g^{\alpha_i \alpha_j}$, где g^{α_i} не является выбранным элементом. Это может значительно ускорить процесс. В итоге построим множество K_2, равное объединению K_1 с выбранными на этом шаге элементами из M_2.

Если ни один элемент не выбран, то $\langle g^A \rangle = \langle K_1 \rangle$. Это также следует из того, что группа $\langle K_1 \rangle$ в этом случае инвариантна относительно действия A.

Далее аналогично строим множества K_r для $r = 3, \ldots$, просматривая и поочередно добавляя элементы из соответствующего множества M_r. Заметим, что на каждом шаге $\langle g^A \rangle = \langle K_r, \bigcup\limits_{i=r+1}^{\infty} M_i \rangle$.

На каком-то шаге получим $K_t = K_{t+1}$, так как группа $\langle g^A \rangle$ по сделанному предположению конечно порождена. Тогда получаем равенство $\langle g^A \rangle = \langle K_t \rangle$. Заметим также, что если подгруппа $\langle g^A \rangle$ имеет s порождающих, то процесс обязательно закончится не позднее чем при рассмотрении K_s. Описанный процесс находит наименьшее по мощности множество порождающих элементов группы $\langle g^A \rangle$. Это позволяет в ряде случаев дать очевидную оценку сверху на число шагов алгоритма. ■

Замечание 1. Если A — коммутативная подгруппа, как далее в протоколе из [2], то рассматриваем не все конечные наборы вида $\alpha_{i_1} \ldots \alpha_{i_t}$, $t = 0, 1, \ldots$, а только те, для которых индексы не убывают.

2. Эффективность метода

Метод нелинейного разложения применим, если в группе G, используемой в качестве платформы шифрования, эффективно решается проблема поиска вхождения элемента в подгруппу, соответствующая классической проблеме вхождения, а также если можно эффективно построить порождающие множества элементов для некоторых конечно порождённых подгрупп, использующихся при шифровании. Пример такого построения указан в лемме 1.

Метод хорошо работает на ряде схем, построенных на полициклических группах, которые сейчас часто предлагаются в качестве платформ [9 – 11]. В полициклических группах любая подгруппа конечно порождена, причем количество её порождающих элементов оценивается сверху длиной полициклического ряда всей группы. Это позволяет дать верхнюю оценку времени работы алгоритма, описанного в лемме 1. В [12] показано, что основные алгоритмические проблемы (построения нормальной формы элемента, определения сопряжённости двух элементов, проблемы вхождения и др.) могут быть решены алгоритмами, работающими в логарифмическом пространстве за квазилинейное время. Более того, если рассматривать постановку проблемы в сжатом виде, причем каждое вхождение слова представлять как strait-line программу, то эти проблемы решаются за полиномиальное время. В [13] использование так называемых TC^0-схем позволило авторам понизить оценки сложности проблем. Наконец, в [14] авторы расширили список алгоритмических проблем для конечно порождённых нильпотентных групп, решаемых алгоритмами с низкой сложностью, включив в него ряд проблем для подгрупп.

3. Примеры криптографического анализа алгебраических схем с использованием метода нелинейного разложения

3.1. Пример 1

Опишем протокол распределения ключа Махалонобиса [2] — аналога классического протокола Масси — Омуры.

Пусть G — группа и $g \in G$. Предположим, что Φ и Ψ — два конечных подмножества группы автоморфизмов $\mathrm{Aut}(G)$, причём элементы из Φ попарно перестановочны с элементами из Ψ. Пусть A и B — подгруппы группы $\mathrm{Aut}(G)$, порождённые множествами Φ и Ψ соответственно.

Алгоритм распределения ключа работает следующим образом:

1) Алиса выбирает случайный автоморфизм $\alpha \in A$, вычисляет g^α и посылает результат Бобу.
2) Боб выбирает случайным образом автоморфизм $\beta \in B$, вычисляет и посылает $(g^\alpha)^\beta$ Алисе.
3) Затем Алиса вычисляет α^{-1} и получает $((g^\alpha)^\beta)^{\alpha^{-1}} = g^\beta$. После этого она выбирает еще один случайный автоморфизм $\gamma \in A$, вычисляет $(g^\beta)^\gamma$ и передаёт его Бобу.
4) Боб находит β^{-1} и получает ключ $K = ((g^\beta)^\gamma)^{\beta^{-1}} = g^\gamma$.
5) Алиса в свою очередь, зная g и γ, тоже вычисляет ключ K.

Криптографический анализ. Заметим, что в алгоритме подгруппы $A, B \leqslant \leqslant \mathrm{Aut}(G)$ конечно порождены.

Предположим, что для группы G эффективно решается проблема поиска вхождения элемента в подгруппу $\langle v^A \rangle$, где $v = (g^\alpha)^\beta$. Пусть известно, что подгруппа $\langle v^A \rangle$ s-порождена для некоторого s.

Замечание 2. В [2] в качестве платформы шифрования предлагается использовать конечно порождённую p-группу, более того, конечную p-группу специального вида. Приведённые условия криптографического анализа в этом случае выполнены. Значение оценочного параметра в оригинальном протоколе очевидно (в общем случае s не больше, чем полициклический ранг группы).

Заметим также, что выше приведён несколько более общий вариант оригинального протокола: автор брал в качестве A и B общую коммутативную подгруппу S группы автоморфизмов $\mathrm{Aut}(G)$. Мы не даём и не используем детали выбора групп и автоморфизмов из [2], так как криптоанализ от них не зависит. Детали только упрощают его реализацию.

Перейдём непосредственно к криптографическому анализу.

По лемме 1 можно эффективно найти порождающее множество элементов для группы $\langle v^A \rangle$. Обозначим эти элементы как $\langle v^{\alpha_1}, \ldots, v^{\alpha_t} \rangle$. Далее отметим, что $(g^\beta)^\gamma \in v^A$, поэтому можно найти его представление вида

$$g^{\beta\cdot\gamma} = V(v^{\alpha_1}, \ldots, v^{\alpha_t}),$$

где $V(x_1, \ldots, x_t)$ — групповое слово от t переменных. Затем в правую часть данного представления вместо $v = (g^\alpha)^\beta$ подставим $g^\alpha = (g^{\alpha\cdot\beta})^{\beta^{-1}}$. Получим

$$V((g^\alpha)^{\alpha_1}, \ldots, (g^\alpha)^{\alpha_t}) = V((v^{\beta^{-1}})^{\alpha_1}, \ldots, (v^{\beta^{-1}})^{\alpha_t}) =$$
$$= V((v^{\alpha_1})^{\beta^{-1}}, \ldots, (v^{\alpha_t})^{\beta^{-1}}) = (V(v^{\alpha_1}, \ldots, v^{\alpha_t}))^{\beta^{-1}} = g^\gamma.$$

Таким образом, не вычисляя ни один из закрытых элементов α, β и γ, можно эффективно найти распределяемый (передаваемый) секретный элемент (ключ или сообщение) g^γ. Это означает, что данная схема является уязвимой.

3.2. Пример 2

В качестве второго примера рассмотрим алгебраический аналог классического протокола Эль-Гамаля, использующий автоморфизмы. Имеется несколько алгебраических версий как самого протокола Эль-Гамаля, так и его обобщения — так называемой системы MOR, относительно которых см. [3, 4].

Алгебраическая версия системы Эль-Гамаля. Соглашения о группе G, элементе $g \in G$ и подгруппах $A, B \leqslant \mathrm{Aut}(G)$ те же, что и в предыдущем примере.

Алгоритм работает следующим образом:

1) Алиса случайным образом выбирает автоморфизм $\alpha \in A$, вычисляет элемент g^α и отправляет его Бобу.
2) Боб хочет послать сообщение m Алисе. Для этого он выбирает случайный автоморфизм $\beta \in B$ и посылает по сети сообщение $(g^\beta, m \cdot g^{\alpha\cdot\beta})$.
3) После этого Алиса вычисляет элемент $(g^\beta)^\alpha = g^{\alpha\cdot\beta}$, находит к нему обратный и получает сообщение $m = m \cdot g^{\alpha\cdot\beta}(g^{\alpha\cdot\beta})^{-1}$.

Криптографический анализ. Пусть для группы G эффективно решается проблема поиска вхождения элемента в подгруппу $\langle g^A\rangle$. Тогда по лемме 1, аналогично примеру 1, найдём порождающее множество для группы $\langle g^A\rangle$, скажем, $\{g^{\alpha_1}, \ldots, g^{\alpha_k}\}$. Из того, что Алиса выбирала автоморфизм $\alpha \in A$, следует, что $g^\alpha \in g^A$, а значит, можно найти представление этого элемента через порождающие:

$$g^\alpha = V(g^{\alpha_1}, \ldots, g^{\alpha_k}).$$

Заменив в правой части g на g^β, получим

$$V((g^\beta)^{\alpha_1}, \ldots, (g^\beta)^{\alpha_k}) = V((g^{\alpha_1})^\beta, \ldots, (g^{\alpha_k})^\beta) = (V(g^{\alpha_1}, \ldots, g^{\alpha_k}))^\beta = g^{\alpha\cdot\beta}.$$

Далее, действуя аналогично Алисе, вычисляем обратный элемент к $g^{\alpha\cdot\beta}$ и расшифровываем сообщение $m = m \cdot g^{\alpha\cdot\beta}(g^{\alpha\cdot\beta})^{-1}$.

Алгебраическая версия системы MOR. Пусть $G = \langle g_1, \ldots, g_n\rangle$ — конечно порождённая группа. Алгоритм распределения ключа работает следующим образом:

1) Алиса выбирает случайный автоморфизм $\varphi \in \mathrm{Aut}(G)$ и случайное натуральное число k, вычисляет значения g_i^φ и $g_i^{\varphi^k}$ для $i = 1, \ldots, n$, которые считаются открытыми. Другими словами, Алиса объявляет открытыми автоморфизмы φ и φ^k. При этом k — закрытый параметр.
2) Боб хочет послать Алисе сообщение m, закодированное как элемент группы G. Для этого Боб выбирает случайное натуральное число l, вычисляет и посылает Алисе набор значений $g_i^{\varphi^l}$ для $i = 1, \ldots, n$, а также элемент $m^{\varphi^{kl}}$. Другими словами, Боб посылает Алисе пару $(\varphi^k, m^{\varphi^{kl}})$.
3) Затем Алиса, зная k, вычисляет $(\varphi^{kl})^{-1}$ и раскрывает сообщение m, подействовав этим автоморфизмом на $m^{\varphi^{kl}}$.

Криптографический анализ. Повторяем рассуждения предыдущего криптографического анализа, позволяющие вычислить по значениям $g_i^{\varphi^k}$ и $g_i^{\varphi^l}$ значение $g_i^{\varphi^{kl}}$ для любого $i = 1, \ldots, n$. Тем самым мы найдём автоморфизм φ^{kl}, вычислим к нему обратный, а затем по элементу $m^{\varphi^{kl}}$ получим сообщение m.

Замечание 3. Для эффективности алгоритма проведённого криптографического анализа необходимо уметь решать проблему поиска вхождения элемента в подгруппу. В [3] в качестве платформы предложено выбирать группу унитреугольных матриц над конечным полем (конечную нильпотентную p-группу, где p — характеристика поля), в которой эта проблема решается легко. В [4] предлагается брать специальную линейную группу над конечным полем, которая, очевидно, сама конечна. Вопрос о сложности алгоритма, решающего в такой группе проблему вхождения элемента в подгруппу, как и вопрос об эффективности построения порождающего множества для подгруппы ещё должен быть изучен. Заметим, что при том и другом предложении для

криптоанализа лучше использовать метод линейного разложения, уже неоднократно упомянутый в данной работе.

Заключение

В работе показано, как метод нелинейного разложения может быть эффективно применен для криптографического анализа ряда схем алгебраической криптографии, использующих автоморфизмы, в частности, для алгебраических аналогов схем Масси — Омуры, Эль-Гамаля и системы MOR. Для его реализации необходимо и достаточно, чтобы в группе, выбранной в качестве платформы шифрования, эффективно решалась проблема вхождения элемента в подгруппу определённого вида, а также можно было бы эффективно построить систему порождающих элементов такой подгруппы. Объяснено, что конечно порождённые и полициклические группы, часто предлагаемые в качестве платформ в современной алгебраической криптографии (в том числе в приведённых в работе примерах криптографического анализа), удовлетворяют этим условиям. Алгоритм построения системы порождающих элементов для таких групп описан в работе.

Выводы. Приведённый криптографический анализ позволяет заключить, что рассматриваемые схемы при определённых естественных и часто предлагаемых условиях оказываются уязвимыми, следовательно, они не могут считаться надёжными.

ЛИТЕРАТУРА

1. *Roman'kov V. A.* A nonlinear decomposition attack // Groups Complexity Cryptology. 2017. V. 8. P. 197–207.
2. *Mahalanobis A.* The Diffie — Hellman key exchange protocol and non-abelian nilpotent groups // Israel J. Math. 2008. V. 165. P. 161–187.
3. *Mahalanobis A.* A simple generalization of El-Gamal cryptosystem to non-abelian groups // Communications in Algebra. 2008. V. 36. P. 3878–3889.
4. *Mahalanobis A.* A simple generalization of El-Gamal cryptosystem to non-abelian groups II // Communications in Algebra. 2012. V. 40. P. 171–186.
5. *Романьков В. А.* Введение в криптографию. Курс лекций. М.: Форум, 2012. 240 с.
6. *Романьков В. А.* Алгебраическая криптография. Омск: Изд-во ОмГУ, 2013. 135 с.
7. *Романьков В. А.* Криптографический анализ некоторых известных схем шифрования, использующих автоморфизмы // Прикладная дискретная математика. 2013. № 3(21). С. 35–51.
8. *Myasnikov A. G. and Roman'kov V. A.* A linear decomposition attack // Groups Complexity Cryptology. 2015. V. 7. P. 81–94.
9. *Eick B. and Kahrobaei D.* Polycyclic groups: A new platform for cryptology? arXiv math.: 0411.077v1 [math.GR].
10. *Gryak K. J. and Kahrobaei D.* The status of polycyclic group-based cryptography: A survey and open problems // Groups Complexity Cryptology. 2017. V. 8. P. 171–186.
11. *Cavallo B. and Kahrobaei D.* A family of polycyclic groups over which the conjugacy problem is NP-complete. arXiv math.: 1403.4153v2 [math. GR], 19 Mar 2014. P. 1–14.
12. *Macdonald J., Miasnikov A., Nikolaev A., and Vassileva S.* Logspace and compressed-word computations in nilpotentgroups. arXiv math.:1503.03888 [math.GR]. 2015.
13. *Macdonald J., Miasnikov A., and Ovchinnikov D.* Low-complexity computations for nilpotent subgroup theorem. arXiv math.: 1706.01092v2 [math. GR] 4 Jul 2017. 23 p.
14. *Miasnikov A. and Weiβ A.* TC^0 circuits for algorithmic problems in nilpotent groups // 42nd Intern. Symp. MFCS. 2017. Article No. 23.

REFERENCES

1. *Roman'kov V. A.* A nonlinear decomposition attack. Groups Complexity Cryptology, 2017, vol. 8, pp. 197–207.
2. *Mahalonobis A.* The Diffie — Hellman key exchange protocol and non-abelian nilpotent groups. Israel J. Math., 2008, vol. 165, pp. 161–187.
3. *Mahalanobis A.* A simple generalization of El-Gamal cryptosystem to non-abelian groups. Communications in Algebra, 2008, vol. 36, pp. 3878–3889.
4. *Mahalanobis A.* A simple generalization of El-Gamal cryptosystem to non-abelian groups II. Communications in Algebra, 2012, vol. 40, pp. 171–186.
5. *Roman'kov V. A.* Vvedenie v kriptografiyu. Kurs lektsiy [Introduction to Cryptography. Lecture Course]. Moscow, Forum Publ., 2012. 240 p. (in Russian)
6. *Roman'kov V. A.* Algebraicheskaya kriptografiya [Algebraic Cryptography]. Omsk, Dostoevsky Omsk State University Publ., 2013. 135 p. (in Russian)
7. *Roman'kov V. A.* Kriptograficheskiy analiz nekotorykh izvestnykh skhem shifrovaniya, ispol'zuyushchikh avtomorfizmy [Cryptographic analysis of some known encryption schemes using automorphisms]. Prikladnaya Diskretnaya Matematika, 2013, no. 3(21), pp. 35–51. (in Russian)
8. *Myasnikov A. G. and Roman'kov V. A.* A linear decomposition attack. Groups Complexity Cryptology, 2015, vol. 7, pp. 81–94.
9. *Eick B. and Kahrobaei D.* Polycyclic groups: A new platform for cryptology? arXiv math.: 0411.077v1 [math.GR].
10. *Gryak K. J. and Kahrobaei D.* The status of polycyclic group-based cryptography: A survey and open problems. Groups Complexity Cryptology, 2017, vol. 8, pp. 171–186.
11. *Cavallo B. and Kahrobaei D.* A family of polycyclic groups over which the conjugacy problem is NP-complete. arXiv math.: 1403.4153v2 [math. GR], 19 Mar 2014, pp. 1–14.
12. *Macdonald J., Miasnikov A., Nikolaev A., and Vassileva S.* Logspace and compressed-word computations in nilpotentgroups. arXiv math.:1503.03888 [math.GR], 2015.
13. *Macdonald J., Miasnikov A., and Ovchinnikov D.* Low-complexity computations for nilpotent subgroup theorem. arXiv math.: 1706.01092v2 [math. GR], 4 Jul 2017. 23 p.
14. *Miasnikov A. and Weiβ A.* TC^0 circuits for algorithmic problems in nilpotent groups. 42nd Intern. Symp. MFCS, 2017. Article no. 23.

ПРИКЛАДНАЯ ТЕОРИЯ ГРАФОВ

УДК 519.1

ПЕРИОДЫ φ-ГРАФОВ

Н. А. Артемова

Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, г. Саратов, Россия

Связный граф с $n \geqslant 3$ вершинами, полученный из контура C_n путём переориентации некоторых его дуг, называется многоугольным графом. Рассмотрим некоторую биекцию φ между множеством стоков и множеством источников многоугольного графа G. Присоединим к G все дуги вида $v\varphi(v)$, где v — сток. Полученный сильносвязный граф будем называть φ-графом. Рассматривая последовательность различных матриц $A, A^2, A^3, \ldots$ (степеней булевой матрицы A), заметим, что эта последовательность конечна. Если A^m — её последний элемент, то $A^{m+1} = A^l$ для некоторого $l \leqslant m$. Число $\mathrm{ind}(A) = l - 1$ называется индексом матрицы A, а число $\mathrm{p}(A) = ((m+1) - l)$ — её периодом. Для графа G с матрицей смежности A положим $\mathrm{ind}(G) = \mathrm{ind}(A)$ и $\mathrm{p}(G) = \mathrm{p}(A)$ (индекс и период графа). Вычислены значения периодов всех неизоморфных φ-графов с числом вершин до 9. Рассчитаны максимальные периоды φ-графов с числом вершин до 17. Доказана теорема, позволяющая вычислить период любого φ-графа. Найдено значение максимального периода n-вершинных φ-графов при чётном n и дана оценка максимального периода при нечётном n.

Ключевые слова: *многоугольный граф, примитивность, φ-граф, индекс графа, период графа.*

DOI 10.17223/20710410/41/5

PERIODS OF φ-GRAPHS

N. A. Artemova

Saratov State University, Saratov, Russia

E-mail: NatalyaKorArt@ya.ru

A connected graph with $n \geqslant 3$ vertices obtained from the circuit C_n by reorienting some of its arcs is called a polygonal graph. We consider a bijection φ between the set of sinks and the set of sources of a polygonal graph G. We attach to G all arcs of type $v\varphi(v)$ where v is a sink. The resulting strongly connected graph is called a φ-graph. When we compute successive powers of a binary Boolean matrix A, the sequence starts to repeat itself at some moment, i.e. we get $A^{m+1} = A^l$ for some $l \leqslant m$. The number $\mathrm{ind}(A) = l - 1$ is called an index, and the value $\mathrm{p}(A) = ((m+1) - l)$ is the period of the matrix A. For the graph G with adjacency matrix A, let $\mathrm{ind}(G) = \mathrm{ind}(A)$ and $\mathrm{p}(G) = \mathrm{p}(A)$ (index and period of the graph). We calculate the values of periods of all not isomorphic φ-graphs with a number of vertices up to nine and the maximal periods of φ-graphs with a number of vertices up to seventeen. We prove the theorem

that allows to compute the period of any φ-graph. Namely, the period of a φ-graph is equal to the greatest common divisor of the lengths of its circuits. The value of the maximal period for n-vertex φ-graph with even n equals $n/2+1$, and the maximal period of a φ-graph with an odd n is less than $\lfloor n/2 \rfloor + 1$. From the theorem for the maximal values of the periods, we obtain some corollaries. Particularly, according to Corollary 1, among the all n-vertex φ-graphs with even n, φ-graphs obtained from the polygonal graphs with one sink and one source have the maximal period.

Keywords: *polygonal graph, primitivity, φ-graph, index and period of graph.*

Введение

Ориентированный граф (далее граф) — это пара $G = (V, \alpha)$, где V — конечное непустое множество, а $\alpha \subseteq V^2$ — бинарное отношение на множестве V. Отношение α называют отношением смежности, а соответствующую ему двоичную булеву матрицу — матрицей смежности графа G. Элементы множества V называются вершинами графа, а пары, входящие в отношение смежности α, — его дугами. Если $(u, v) \in \alpha$, то говорят, что вершина u является началом дуги (u, v), а вершина v — её концом. При $u = v$ получается петля (u, u). Считаем, что каждая вершина графа инцидентна некоторой дуге, т. е. является началом или концом некоторой дуги.

Говорят, что вершина v достижима из вершины u за $k \geqslant 1$ шагов, если существует последовательность примыкающих дуг (маршрут) $(w_0, w_1), (w_1, w_2), \ldots, (w_{k-1}, w_k)$, где $w_0 = u$ и $w_k = v$. Если A — матрица смежности орграфа G, то последнее определение означает, что на пересечении строки, соответствующей элементу u, и столбца, соответствующего элементу v, в матрице-степени A^k стоит 1.

Маршрут с неповторяющимися вершинами $C_n = v_1 v_2 \ldots v_n v_1$, в котором начало и конец совпадают, называется n-элементным контуром.

Граф $G = (V, \alpha)$ по определению является функциональным, если его отношение смежности функционально (т. е. из каждой вершины исходит точно одна дуга).

Функциональный граф называется связным, если он содержит точно один контур. Под высотой вершины в функциональном графе понимается расстояние от неё до контура, т. е. минимальная из длин цепей с началом в данной вершине и концом в вершине, принадлежащей контуру.

Под конечной динамической системой понимается пара (S, δ), где S — конечное непустое множество состояний системы, $\delta : S \to S$ — отображение множества состояний в себя, называемое эволюционной функцией системы. Таким образом, каждой конечной динамической системе сопоставляется карта, представляющая собой граф с множеством вершин S и дугами, проведёнными из каждой вершины $s \in S$ в вершину $\delta(s)$. Этот граф является функциональным. Компоненты связности графа, задающего динамическую систему, называются её бассейнами. Получается, что каждый бассейн представляет собой контур с входящими в него деревьями. Контуры, в свою очередь, называются предельными циклами, или аттракторами. Индексом состояния называют его расстояние до аттрактора, а периодом — длину принимающего аттрактора.

Связный граф с $n \geqslant 3$ вершинами, полученный из контура C_n путем переориентации некоторых его дуг, называется многоугольным графом.

Граф называется примитивным, если существует целое число $r \geqslant 1$, такое, что каждая вершина графа достижима из любой вершины за r шагов (иначе говоря, если в матрице A^r все элементы равны 1). Таким образом, каждый примитивный граф

является сильносвязным (любые две вершины взаимно достижимы), но обратное не верно.

1. Об индексах и периодах

Пусть A — булева матрица. Рассматривая последовательность различных матриц $A, A^2, A^3, \ldots$, заметим, что эта последовательность конечна и что, если A^m — её последний элемент, то $A^{m+1} = A^l$ для некоторого $l \leqslant m$. Число $\mathrm{ind}(A) = l - 1$ называется индексом матрицы A, а число $\mathrm{p}(A) = ((m+1) - l)$ — её периодом. Так определённые индекс и период матрицы A — это её индекс и период в динамической системе булевых матриц соответствующей размерности с эволюционной функцией $\delta(A^k) = A^{k+1}$. Для графа G с матрицей смежности A положим $\mathrm{ind}(G) = \mathrm{ind}(A)$ и $\mathrm{p}(G) = \mathrm{p}(A)$ (индекс и период графа). Пару $t(G) = (\mathrm{ind}(G), \mathrm{p}(G))$ назовём типом графа G. Об индексах и периодах графов см. [1].

Ряд работ посвящён двоичным булевым матрицам с минимально возможным типом $(0, 1)$ (идемпотентные матрицы) [2]. Индексы и периоды относятся к числу важнейших параметров, связываемых с графами. Решению проблем, связанных с этими параметрами, посвящены работы [1 – 6]. Об индексах состояний в динамических системах, связанных с графами, см. [7 – 11]. Не для всех графов индексы и периоды аналитически вычислены. Известны, например, следующие результаты.

Теорема 1 [3]. Индекс функционального графа равен уменьшенной на единицу максимальной из высот его элементов, а период — наименьшему общему кратному длин его контуров.

Теорема 2 [3]. Бесконтурный граф имеет индекс, равный максимальной из длин его цепей, и период, равный 1.

Каковы максимальные значения индекса и периода для n-вершинного графа? Для периода точная формула неизвестна, асимптотической оценкой является $(n \ln n)^{1/2}$ [6]. Что касается индекса, то имеются компьютерные вычисления [4], которые показывают, что для графа с n вершинами справедливо неравенство $\mathrm{ind} \leqslant (n-1)^2$.

2. О периодах φ-графов

Вершина графа называется источником, если в неё не входит ни одна дуга, и стоком, если из неё не исходит ни одна дуга. Количество источников в многоугольном графе равно количеству стоков. Пусть φ — некоторая биекция между множеством стоков и множеством источников данного многоугольного графа G. Если к G присоединить все дуги вида $v\varphi(v)$, где v — сток, получится сильносвязный граф, назовём его φ-графом.

Конструкция φ-графа предложена в [12] в связи с проблемой описания минимальных примитивных расширений для многоугольных графов. В частности, доказано, что любой φ-граф, полученный из данного многоугольного графа, является его минимальным сильносвязным расширением.

В каждом многоугольном графе есть по крайней мере один источник. Такая вершина имеет степень исхода 2. Эту степень она сохранит и в любом φ-графе, связанном с исходным многоугольным графом. Следовательно, φ-графы не являются функциональными графами и к ним неприменима теорема 1. Будучи сильносвязными, φ-графы не удовлетворяют и условию теоремы 2, так что вопрос об индексах φ-графов требует отдельного рассмотрения.

Был проведен вычислительный эксперимент, в ходе которого были найдены все неизоморфные φ-графы размерности от 3 до 9 вершин и вычислены их периоды. Получены следующие результаты:

— $n = 3$. Существует один φ-граф с 3 вершинами. Он имеет период 1.
— $n = 4$. Есть всего три неизоморфных φ-графа с 4 вершинами. Два графа имеют период 2 и один — период 3.
— $n = 5$. Есть всего четыре неизоморфных φ-графа с 5 вершинами. Данные графы имеют период 1.
— $n = 6$. Есть всего одиннадцать неизоморфных φ-графов с 6 вершинами. Три графа имеют период 1, семь графов имеют период 2 и один граф — период 4.
— $n = 7$. Есть всего девятнадцать неизоморфных φ-графов с 7 вершинами. Восемнадцать графов имеют период 1 и один граф — период 3.
— $n = 8$. Есть всего сорок семь неизоморфных φ-графов с 8 вершинами. Двадцать два графа имеют период 1, двадцать один граф — период 2, два графа — период 3 и по одному графу имеют периоды 4 и 5.
— $n = 9$. Есть всего сто четырнадцать неизоморфных φ-графов с 9 вершинами. Сто тринадцать графов имеют период 1 и один граф имеет период 3.

Вычислены максимальные периоды графов каждой размерности до 17 вершин (таблица).

Максимальные периоды φ-графов

Размерность	Максимальный период
3	1
4	3
5	1
6	4
7	3
8	5
9	3
10	6
11	5
12	7
13	5
14	8
15	7
16	9
17	7

На основе полученных данных сформулировано следующее утверждение.

Теорема 3. Период φ-графа равен наибольшему общему делителю длин его контуров.

Доказательство. Пусть дан φ-граф G с матрицей смежности A; $C_1, C_2, \ldots, C_t$ — множество всех контуров графа G; $l_1, l_2, \ldots, l_t$ — соответствующие длины контуров.

1) Если граф G примитивен, то по определению существует целое число $r \geqslant 1$, такое, что в матрице A^r все элементы равны 1, т. е. $A^r = A^{r+1}$. По определению период графа G равен $(r + 1) - r$, т. е. $\mathrm{p}(G) = 1$. По критерию примитивности наибольший общий делитель длин всех контуров примитивного графа равен 1. Таким образом, $\mathrm{p}(G) = 1$, $(l_1, l_2, \ldots, l_t) = 1$. Следовательно, $\mathrm{p}(G) = (l_1, l_2, \ldots, l_t)$.

2) Пусть граф G непримитивен, $\mathrm{p}(G) \neq 1$ и $(l_1, l_2, \ldots, l_t) \neq 1$. Будем обозначать через a_{ij} элемент матрицы смежности A, находящийся на пересечении строки i и столбца j. Элемент a_{ij}^k — аналогичный элемент в матрице-степени A^k.

Если элемент $a_{ij}^k = 1$, то это означает, что в исходной матрице A вершина j достижима из вершины i за k шагов (существует путь длины k из вершины i в вершину j). Если в матрице A^{k_1} элемент $a_{ij}^{k_1}$ равен 1 и в матрице A^{k_2} элемент $a_{ij}^{k_2}$ равен 1 $(k_1 < k_2)$, то в матрице A существуют пути длины k_1 и k_2, соединяющие вершины i и j. Так как в φ-графе любые две вершины взаимно достижимы, то длину простого пути из вершины i в вершину j обозначим $k = \min\{k_s : a_{ij}^{k_s} = 1, s = 1, 2, \ldots\}$.

Рассмотрим некоторый путь длины k_s из вершины i в вершину j, где $s = 1, 2, \ldots$ Исходя из структуры φ-графа (в любом φ-графе каждая дуга принадлежит некоторому контуру), имеем, что длина пути из вершины i в вершину j может быть увеличена только за счёт прохождения по некоторому контуру. Таким образом, $k_s = k + (x_1^s l_1 + x_2^s l_2 + \ldots + x_t^s l_t)$, где $x_1^s, x_2^s, \ldots, x_t^s$ — целые неотрицательные числа.

Так как граф G непримитивен, $A^{m+1} = A^l$ для некоторого $l < m$. Для любых вершин i и j графа G, таких, что $a_{ij}^l = a_{ij}^{m+1} = 1$, имеем $l = k + (x_1^l l_1 + x_2^l l_2 + \ldots + x_t^l l_t)$, $(m+1) = k + (x_1^{m+1} l_1 + x_2^{m+1} l_2 + \ldots + x_t^{m+1} l_t)$.

Период $\mathrm{p}(A)$ равен $(m+1) - l$, следовательно, $m + 1 = \mathrm{p}(A) + l$. Таким образом,

$$\mathrm{p}(A) + l = k + (x_1^{m+1} l_1 + x_2^{m+1} l_2 + \ldots + x_t^{m+1} l_t),$$
$$l = k + (x_1^l l_1 + x_2^l l_2 + \ldots + x_t^l l_t) = k + (x_1^{m+1} l_1 + x_2^{m+1} l_2 + \ldots + x_t^{m+1} l_t) - \mathrm{p}(A),$$
$$\mathrm{p}(A) = k + (x_1^{m+1} l_1 + x_2^{m+1} l_2 + \ldots + x_t^{m+1} l_t) - (k + (x_1^l l_1 + x_2^l l_2 + \ldots + x_t^l l_t)),$$
$$\mathrm{p}(A) = (x_1^{m+1} l_1 + x_2^{m+1} l_2 + \ldots + x_t^{m+1} l_t) - (x_1^l l_1 + x_2^l l_2 + \ldots + x_t^l l_t) = (z_1^p l_1 + z_2^p l_2 + \ldots + z_t^p l_t),$$

где $z_i^p = (x_i^{m+1} - x_i^l)$; x_i — целые неотрицательные числа; z_i — целые числа, $i = 1, 2, \ldots, t$.

Так как $A^l = A^{m+1}$, $A^l = A^{\mathrm{p}(A)+l} = A^{2\mathrm{p}(A)+l} = A^{3\mathrm{p}(A)+l} = \ldots$, то $A^l = A^{\alpha \mathrm{p}(A)+l}$, $\alpha = 1, 2, \ldots$ Следовательно, $\alpha \mathrm{p}(A) = (z_1 l_1 + z_2 l_2 + \ldots + z_t l_t)$, $\alpha = 1, 2, \ldots$

При $z_1 = 1$, $z_2 = z_3 = \ldots = z_t = 0$ имеем $\alpha_1 \mathrm{p}(A) = l_1$. Аналогично $\alpha_s \mathrm{p}(A) = l_s$, $s = 1, 2, \ldots, t$, т. е. длина любого контура графа G представима в виде $\alpha_i p$, где $p = \mathrm{p}(A)$. Следовательно, период графа G является общим делителем длин его контуров.

Очевидно, что $p \leqslant \min\{l_i : i = 1, 2, \ldots, t\}$. Докажем, что p — наибольший делитель. При $p = \min\{l_i : i = 1, 2, \ldots, t\}$ это очевидно.

Рассмотрим случай, когда $p < \min\{l_i : i = 1, 2, \ldots, t\}$, и докажем, что p — наибольший делитель длин контуров графа G.

От противного. Пусть p — общий делитель длин контуров, но не наибольший. Тогда граф G состоит из контуров вида $l_i = \alpha_i p$, где $\alpha_i = a c_i$, a, c_i — целые числа, $i = 1, 2, \ldots, t$ (т. е. наибольший общий делитель равен ap). Имеем

$$p = (z_1 l_1 + z_2 l_2 + \ldots + z_t l_t) = (z_1 a p c_1 + z_2 a p c_2 + \ldots + z_t a p c_t),$$
$$p = ap(z_1 c_1 + z_2 c_2 + \ldots + z_t c_t),$$
$$1/a = z_1 c_1 + z_2 c_2 + \ldots + z_t c_t,$$

где p, a, c_i, z_i — целые числа. Последнее уравнение имеет решение в целых числах только при $a = 1$. Следовательно, p — наибольший общий делитель длин контуров графа G. ■

Теорема 4. Максимальный период φ-графа с числом вершин n не превышает $\lfloor n/2 \rfloor + 1$, причём эта оценка достигается при чётном n.

Доказательство. n-Вершинный φ-граф состоит из $n+k$ дуг (k — количество стоков/источников в исходном многоугольном графе), а сумма длин его контуров равна $n+2k$. По теореме 3 $p = \mathrm{p}(G) = (l_1, l_2, \ldots, l_t)$, $p \leqslant \min\{l_i\}$, где l_i — длины контуров графа. Очевидно, что максимальный период $p = \min\{l_i\}$ имеют графы, у которых контур минимальной длины максимален.

При чётном n длина минимального контура будет максимальной при $k=1$, $t=2$, $l_1 = l_2 = (n+2k)/2 = n/2 + k = n/2 + 1$, т. е. граф имеет два контура длины $n/2+1$. Следовательно, $p = n/2+1$.

При нечётном n гипотетически максимальная длина минимального контура достигается при $k=1$, $t=2$, $l_1 = \lfloor n/2 \rfloor + 1$, $l_2 = \lfloor n/2 \rfloor + 2$ и равна l_1. Так как $(\lfloor n/2 \rfloor + 1, \lfloor n/2 \rfloor + 2) = 1$, перебрав все возможные длины контуров при $k = 1, 2, \ldots$, получаем, что $\mathrm{p}(G) < \lfloor n/2 \rfloor + 1$. ■

φ-Графы, полученные из многоугольных графов с одним источником и стоком, будем называть φ_1-графами. Такой граф имеет два контура. Из теоремы 4 получаем три следствия.

Следствие 1. Максимальный период среди всех n-вершинных φ-графов при чётном n имеют φ_1-графы.

Следствие 2. Максимальный период n-вершинных φ-графов при чётном n равен $n/2+1$.

Следствие 3. Максимальный период n-вершинных φ-графов при нечётном n имеет оценку $\mathrm{p}(G) < \lfloor n/2 \rfloor + 1$.

3. Пояснения

Коэффициенты $z_i = x_i^{'} - x_i^{''}$ — целые числа. Эти коэффициенты могут принимать отрицательные значения (при $x_i^{''} > x_i^{'}$). Как показано в доказательстве теоремы 3, $p \leqslant \min\{l_i : i = 1, 2, \ldots, t\}$. Случаи $p = \min\{l_i : i = 1, 2, \ldots, t\}$ и $p = 1$ не нуждаются в пояснениях.

Рассмотрим случай $p < \min\{l_i : i = 1, 2, \ldots, t\}$ на примере конкретного φ-графа G (рис. 1) с матрицей смежности A, которая имеет следующий вид:

$$A = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}.$$

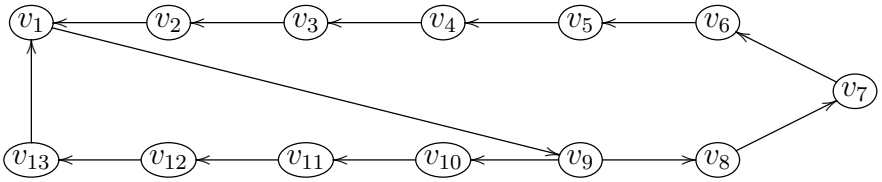

Рис. 1. φ-Граф G с матрицей смежности A

Для данного графа $l = 19$, $\mathrm{ind}(G) = 18$, $\mathrm{p}(G) = 3$, $l_1 = 6$, $l_2 = 9$.

Рассмотрим пути из вершины v_1 в вершину v_3. Элемент $a^k_{1,3}$ матриц A^k равен 1 при $k = 7, 13, 16, 19, 22, \ldots$ Положим $k = 7$, $k_1 = 13$, $k_2 = 16$, $k_3 = 19$, $k_4 = 22$, $\ldots$ Запишем, согласно формулам из теоремы 3, $k_1 = k + x^1_1 l_1 + x^1_2 l_2$. Подставим известные значения k, k_1, l_1, l_2 и вычислим значения коэффициентов x^1_1, x^1_2. Получим $x^1_1 = 1$, $x^1_2 = 0$.

Аналогичные действия выполним и для k_2, k_3, k_4:

$$\begin{aligned} k_2 &= k + x^2_1 l_1 + x^2_2 l_2, \quad x^2_1 = 0, \quad x^2_2 = 1, \\ k_3 &= k + x^3_1 l_1 + x^3_2 l_2, \quad x^3_1 = 2, \quad x^3_2 = 0, \\ k_4 &= k + x^4_1 l_1 + x^4_2 l_2, \quad x^4_1 = 1, \quad x^4_2 = 1. \end{aligned}$$

Так как $l = k_3 = k + x^3_1 l_1 + x^3_2 l_2$ и $k_4 = k + x^4_1 l_1 + x^4_2 l_2 = l + p = k_3 + p = k + x^3_1 l_1 + x^3_2 l_2 + p$, то $p = k + x^4_1 l_1 + x^4_2 l_2 - (k + x^3_1 l_1 + x^3_2 l_2) = (x^4_1 - x^3_1) l_1 + (x^4_2 - x^3_2) l_2$, где $x^4_1 = 1$, $x^4_2 = 1$, $x^3_1 = 2$, $x^3_2 = 0$. Получаем $p = (1-2)l_1 + (1-0)l_2 = -l_1 + l_2 = -6 + 9 = 3$. В обозначениях доказательства теоремы 3 имеем $z_1 = (x^4_1 - x^3_1) = -1$ — отрицательный коэффициент.

Заключение

Доказана теорема, позволяющая вычислить период любого φ-графа. Дана оценка максимального периода n-вершинных φ-графов и показана её корректность.

ЛИТЕРАТУРА

1. *Салий В. Н.* Отказоустойчивость и оптимизация дискретных систем с заданными индексом и периодом // Вестник Томского государственного университета. Приложение. 2006. № 17. С. 222–225.
2. *Chaudhuri R. and Mukherdjea A.* Idempotent Boolean matrices // Semigroup Forum. 1980. V. 21. P. 273–282.
3. *Максимов А. А., Салий В. Н.* Индексы и периоды нечетких матриц и графов // Теоретические проблемы информатики и её приложения. Саратов: Изд-во Сарат. ун-та, 2006. Вып. 7. С. 87–95.
4. *Максимов А. А.* Об индексе и периоде нечеткой матрицы. Саратов, 2005. Деп. в ВИНИТИ 20.01.05. № 78-В2005. 11 с.
5. *Бар-Гнар Р. И., Фомичев В. М.* О минимальных примитивных матрицах // Прикладная дискретная математика. Приложение. 2014. № 7. С. 7–9.
6. *Miller W.* The maximum order of an element of a finite symmetric group // Amer. Math. Monthly. 1987. No. 94. P. 497–506.
7. *Barbosa V. C.* An Atlas of Edge-Reversal Dynamics. London: Chapman&Hall/CRC, 2001. 385 p.
8. *Салий В. Н.* Об одном классе конечных динамических систем // Вестник Томского государственного университета. Приложение. 2005. № 14. С. 23–26.

9. *Власова А. В.* Индексы в динамической системе (B, δ) двоичных векторов // Изв. Сарат. ун-та. Нов. сер. Сер. Математика. Механика. Информатика. 2011. Т. 11. Вып. 3. С. 116–122.
10. *Жаркова А. В.* Индексы в динамической системе двоичных векторов, ассоциированных с ориентациями циклов // Прикладная дискретная математика. 2012. № 2. С. 79–85.
11. *Жаркова А. В.* Индексы состояний в динамической системе двоичных векторов, ассоциированных с ориентациями пальм // Изв. Сарат. ун-та. Нов. сер. Сер. Математика. Механика. Информатика. 2016. Т. 16. Вып. 4. С. 475–484.
12. *Салий В. Н.* Минимальные примитивные расширения ориентированных графов // Прикладная дискретная математика. 2008. № 1. С. 116–119.

REFERENCES

1. *Salii V. N.* Otkazoustoychivost' i optimizatsiya diskretnykh sistem s zadannymi indeksom i periodom [Fault tolerance and optimization of discrete systems with specified index and period]. Vestnik TSU. Prilozhenie, 2006, no. 17, pp. 222–225. (in Russian)
2. *Chaudhuri R. and Mukherdjea A.* Idempotent Boolean matrices. Semigroup Forum, 1980, vol. 21, pp. 273–282.
3. *Maximov A. A. and Salii V. N.* Indeksy i periody nechetkikh matrits i grafov [Indices and periods of fuzzy matrices and graphs]. Theoretical Problems of Computer Science and its Applications. Saratov, Saratov Univ. Press, 2006, vol. 7, pp. 87–95. (in Russian)
4. *Maximov A. A.* Ob indekse i periode nechetkoy matritsy [On the Index and Period of a Fuzzy Matrix]. Saratov, 2005. Dep. v VINITI 20.01.05, no. 78-B2005, 11 p. (in Russian)
5. *Bar-Gnar R. I. and Fomichev V. M.* O minimal'nykh primitivnykh matritsakh [On minimal primitive matrices]. Prikladnaya Diskretnaya Matematika. Prilozhenie, 2014, no. 7, pp. 7–9. (in Russian)
6. *Miller W.* The maximum order of an element of a finite symmetric group. Amer. Math. Monthly, 1987, no. 94, pp. 497–506.
7. *Barbosa V. C.* An Atlas of Edge-Reversal Dynamics. London, Chapman&Hall/CRC, 2001. 385 p.
8. *Salii V. N.* Ob odnom klasse konechnykh dinamicheskikh sistem [A class of finite dynamical systems]. Vestnik TSU. Prilozhenie, 2005, no. 14, pp. 23–26. (in Russian)
9. *Vlasova A. V.* Indeksy v dinamicheskoy sisteme (B, δ) dvoichnykh vektorov [Indices in dynamical system (B, δ) of binary vectors]. Izv. Saratov Univ. (N. S.), Ser. Math. Mech. Inform., 2011, vol. 11, iss 3, pt. 1, pp. 116–122. (in Russian)
10. *Zharkova A. V.* Indeksy v dinamicheskoy sisteme dvoichnykh vektorov, assotsiirovannykh s oriyentatsiyami tsiklov [Indices in dynamic system of binary vectors associated with cycles orientations]. Prikladnaya Diskretnaya Matematika, 2012, no. 2(16), pp. 79–85. (in Russian)
11. *Zharkova A. V.* Indeksy sostoyaniy v dinamicheskoy sisteme dvoichnykh vektorov, assotsiirovannykh s oriyentatsiyami pal'm [Indices of states in dynamical system of binary vectors associated with palms orientations]. Izv. Saratov Univ. (N. S.), Ser. Math. Mech. Inform., 2016, vol. 16, iss 4, pp. 475–484. (in Russian)
12. *Salii V. N.* Minimal'nyye primitivnyye rasshireniya oriyentirovannykh grafov [Minimal primitive extensions of oriented graphs]. Prikladnaya Diskretnaya Matematika, 2008, no. 1, pp. 116–119. (in Russian)

УДК 004.056

ПРОВЕРКА СООТВЕТСТВИЯ ОРИЕНТИРОВАННОГО ГРАФА АЛГЕБРАИЧЕСКОЙ РЕШЁТКЕ

С. В. Белим, Н. Ф. Богаченко

Омский государственный университет им. Ф. М. Достоевского, г. Омск, Россия

Рассмотрена проблема проверки изоморфности ориентированного графа диаграмме некоторой решётки. Исследованы три типа конечных решёток, используемых в моделях разграничения доступа. Построен алгоритм проверки соответствия ориентированного графа прямому произведению решётки подмножеств и линейной решётки. Алгоритм основан на анализе структуры двух множеств: множества вершин, покрывающих ровно одну вершину, и множества атомарных вершин. Доказано, что вычислительная сложность алгоритма проверки не превосходит $O(n^3)$.

Ключевые слова: *граф, теория решёток, мандатное разграничение доступа.*

DOI 10.17223/20710410/41/6

THE CHECK OF THE CORRESPONDENCE OF THE DIRECTED GRAPH TO THE ALGEBRAIC LATTICE

S. V. Belim, N. F. Bogachenko

Dostoevsky Omsk State University, Omsk, Russia

E-mail: belimsv@omsu.ru, nfbogachenko@mail.ru

The isomorphism check problem for a directed graph and a lattice diagram is considered in this paper. Three specific finite lattices used in the access control models are investigated. Special attention is paid to MLS-lattice which is the Cartesian product of the lattice of subsets and the linear lattice. Necessary and sufficient conditions for the isomorphism of a finite lattice S to the MLS-lattice are found. For the lattice S, these conditions define some limitations to the number of all nodes, of atomic nodes, and of nodes covered by each element of the lattice S. An algorithm which checks the correspondence of the directed graph to the MLS-lattice is offered. This algorithm is based on the structure of two sets: the set of the nodes which cover exactly one node and the set of the atomic nodes. The following conditions are checked: the number of nodes of the directed graph $n = 2^{n_1}n_2$; all nodes of the directed graph cover at least two others except the nodes $x_1, \ldots, x_{n_1}$, t, $t_1, \ldots, t_{n_2-1}$, wherein t is an outlet of the directed graph; the nodes $x_1, \ldots, x_{n_1}, t_1, \ldots, t_{n_2-1}$ cover exactly one node; the nodes $x_1, \ldots, x_{n_1}, t_1$ are atomic nodes; the nodes $t_{n_2-1}, \ldots, t_1, t$ form a chain $t_{n_2-1} \succ \ldots \succ t_1 \succ t$ in which the nodes successively cover each other. We prove that the computing complexity of this algorithm does not exceed $O(n^3)$.

Keywords: *graph, lattice theory, mandatory access control.*

Введение

Работа посвящена проблеме поиска решётки, соответствующей заданному орграфу G. Данная проблема находит применение при построении и оптимизации структур, на которых строится политика безопасности компьютерных систем [1]. Можно выделить два случая:

1) Орграф G может быть дополнен до диаграммы некоторой решётки без добавления новых вершин. Необходимо определить соответствующую ему решётку. При этом под дополнением графа понимается построение его транзитивного замыкания — пополнение графа всеми дугами, соединяющими достижимые вершины.
2) Для построения диаграммы некоторой решётки необходимо дополнить орграф G как вершинами, так и рёбрами. В этом случае необходимо определить минимальную решётку, диаграмма которой включает данный орграф. Минимизируется число элементов решётки.

Остановимся на первой задаче, которая частично связана с пятой проблемой Биркгофа [2]. Различные вопросы, касающиеся графов, соответствующих решёткам, рассматриваются в [3 – 7]. Во всех этих работах авторы исходят из существования некоторой решётки, строят соответствующей ей граф и проводят его исследования. Актуальной представляется и обратная задача — исследовав свойства графа, сопоставить ему некоторую решётку. Нас интересует решётка, полученная в результате прямого произведения решётки подмножеств и линейной решётки. Основная цель исследования — построение и оценка вычислительной сложности алгоритма проверки соответствия ориентированного графа такой решётке.

1. Решёточный граф

Определение 1. Частично упорядоченное множество S называется решёткой, если для любых двух его элементов $x, y \in S$ в множестве S можно найти их точную верхнюю и точную нижнюю грани: $\sup\{x, y\}, \inf\{x, y\} \in S$.

Орграф $G = (V, E)$, не содержащий ориентированных циклов, задаёт отношение частичного порядка на множестве своих вершин: $v_1 \geqslant v_2$ $(v_1, v_2 \in V)$ тогда и только тогда, когда в орграфе существует ориентированный путь $p(v_1, v_2)$. Очевидно, что полученное множество не всегда является решёткой.

Определение 2. Ориентированный граф будем называть решёточным, если его вершины образуют решётку с отношением порядка «$\geqslant$», порождённым ориентированными путями орграфа.

Имея в распоряжении матрицу смежности орграфа, несложно доказать следующий факт.

Теорема 1. Трудоёмкость проверки, является ли орграф решёточным, не превышает $\mathrm{O}(n^4)$.

Рассмотрим далее некоторые частные решения задачи восстановления решётки. На сегодняшний день в моделях безопасности компьютерных систем получили широкое распространение три решётки:

1) Линейная решётка. Под линейно упорядоченным множеством B понимается множество, для любых двух элементов которого определено отношение порядка. Как легко показать, такое множество образует решётку: если $a \geqslant b$, то $\sup\{a, b\} = a$ и $\inf\{a, b\} = b$. Решётка данного вида является наиболее распространённой в системах защиты информации, она описывает уровни доступа к данным. Если $B = \{0, 1, \ldots, m-1\}$, то линейная решётка обозначается $SL(m)$.
2) Решётка подмножеств. Пусть задано множество A мощности n. Рассмотрим множество всех его подмножеств $2^A = \{X : X \subseteq A\}$. Введём на множестве 2^A отношение порядка: $\forall X_1, X_2 \in 2^A\,(X_1 \geqslant X_2 \Leftrightarrow X_1 \supseteq X_2)$. Легко доказать, что

отношение $X_1 \supseteq X_2$ есть отношение нестрогого порядка, а множество 2^A частично упорядоченное: так, не для любой пары элементов определено отношение порядка. Наибольшую нижнюю и наименьшую верхнюю грани определим на основе операций пересечения и объединения множеств: $\inf\{X_1, X_2\} = X_1 \cap X_2$, $\sup\{X_1, X_2\} = X_1 \cup X_2$. Также нетрудно доказать, что $(2^A, \supseteq, \cup, \cap)$ — решётка. Примерами её использования в реальных компьютерных системах может служить множество атрибутов файла и зависимость выходной величины от подмножества множества входных элементов. Если $A = \{1, 2, \ldots, n\}$, то решётка подмножеств обозначается $SX(n)$ и содержит 2^n элементов.

3) MLS-решётка — прямое (декартово) произведение первых двух решёток.

Алгебраическая решётка используется в моделях безопасности для задания уровней секретности информации и уровней доверия к пользователям и называется решёткой ценностей. Рассмотрим задачу проверки соответствия указанным решёткам заданного решёточного графа. Для этого потребуется ряд определений и теорема об изоморфности MLS-решётке.

2. Исследование MLS-решётки

Из определения решётки нетрудно вывести, что любое конечное подмножество решётки имеет точную верхнюю и нижнюю грани. В частности, точную верхнюю и нижнюю грань имеет любое подмножество конечной решётки S (в частности, вся S). Обозначим $\sup S = I$, $\inf S = O$ — таким образом, любой элемент конечной решётки S меньше или равен I и больше или равен O.

Определение 3. Будем говорить, что элемент y решётки S покрывает элемент x этой решётки (обозначается $y \succ x$), если $y > x$ и не существует такого $z \in S$, что $y > z > x$.

Интервалом $[x, y]$ для элементов $x, y \in S$, $x \leqslant y$, называется множество таких элементов $z \in S$, что $x \leqslant z \leqslant y$. Таким образом, y покрывает x в точности тогда, когда интервал $[x, y]$ содержит лишь сами элементы x и y.

Лемма 1. Пусть S — конечная решётка, $x, y \in S$, $x < y$. Тогда найдётся такой $z \in [x, y]$, что $z \succ x$.

Доказательство. Пусть y не покрывает x. Тогда по определению можно построить такой y_1, что $x < y_1 < y$. По построению $y_1 \in [x, y]$. Если y_1 также не покрывает x, можем аналогично построить y_2, такой, что $x < y_2 < y_1$. При этом, поскольку $y_1 < y$, y_2 также лежит в интервале $[x, y]$. Продолжая аналогичные построения, получаем цепочку $y_1 > y_2 > y_3 > \ldots$ различных элементов, лежащих в интервале $[x, y]$.

Но поскольку решётка S конечна, в интервале $[x, y]$ может лежать лишь конечное число различных элементов, то есть в какой-то момент мы не сможем построить очередной y_i. Это означает, что y_{i-1} покрывает x. ■

Заменив в доказательстве знаки всех неравенств, можно сформулировать двойственное утверждение: в любом интервале $[x, y]$ конечной решётки S, $x < y$, найдётся такой $z \in [x, y]$, что $z \prec y$.

Определение 4. Элемент $a \in S$ называется атомарным элементом решётки S, имеющей инфинум $O = \inf S$, если $a \succ O$.

Отметим, что атомарные элементы не могут покрывать отличных от O элементов решётки, так как если атомарный элемент a покрывает какой-то отличный от O элемент b, то $a > b > O$ и a не покрывает O.

Определение 5. Пусть S, S' — решётки. Отображение $\psi : S \to S'$ называется порядковым гомоморфизмом решёток, или изотонным отображением, если $x \leqslant y$ влечёт за собой $\psi(x) \leqslant \psi(y)$ для любых $x, y \in S$.

Таким образом, изотонное отображение сохраняет порядок элементов решётки.

Определение 6. Прямым произведением решёток A и B называется решётка $S = A \times B$, определённая на декартовом произведении множеств A и B следующим образом. Для $a, a' \in A$, $b, b' \in B$ полагаем $(a, b) \leqslant (a', b')$ в точности тогда, когда $a \leqslant a'$ и $b \leqslant b'$.

При этом супремумом и инфинумом элементов $(a, b), (a', b') \in A \times B$ будут соответственно $(\sup\{a, a'\}, \sup\{b, b'\})$ и $(\inf\{a, a'\}, \inf\{b, b'\})$. Если решётки A и B имеют нижние грани, то $\inf(A \times B) = (\inf A, \inf B)$.

Отметим, что элемент (a, b) решётки $A \times B$ покрывает элемент (a', b') той же решётки в том и только в том случае, когда

$$a \succ a',\ b = b' \quad \text{или} \quad a = a',\ b \succ b'. \tag{1}$$

Действительно, если выполнена, например, первая пара условий, то $(a, b) > (a', b')$ и если найдётся такой $(x, y) \in A \times B$, что $(a, b) > (x, y) > (a', b')$, то $a > x > a'$, что противоречит условию $a \succ a'$. Обратно, предположим, что для $(a, b) > (a', b')$ условие (1) не выполняется. Тогда либо $a > a'$, $b > b'$, либо $b = b'$, но a не покрывает a', либо $a = a'$, но b не покрывает b'. В первом случае $(a, b) \geqslant (a, b') > (a', b')$, во втором случае найдётся такой $x \in A$, что $a > x > a'$, и тогда $(a, b) > (x, b) > (a', b')$, аналогично в третьем случае $(a, b) > (a, y) > (a', b')$ для такого $y \in B$, что $b > y > b'$. Так или иначе, получаем, что при невыполнении условия (1) элемент (a, b) решётки $A \times B$ не может покрывать её элемент (a', b').

Определение 7. Решётки S и S' называются изоморфными, если существует биективное изотонное отображение $\psi : S \to S'$. Это отображение называется изоморфизмом решёток S и S'.

Любая линейная решётка B при $|B| = m$ изоморфна решётке $SL(m)$. Любая решётка 2^A при $|A| = n$ изоморфна решётке $SX(n)$. Отметим также, что если решётка A изоморфна A', а B изоморфна B', то и $A \times B$ изоморфна $A' \times B'$ — их изоморфизм получается как прямое произведение изоморфизмов решёток-сомножителей.

Напомним, что цепью в частично упорядоченном множестве S называется его вполне упорядоченное подмножество; так, конечная цепь длины m представляет собой набор элементов $b_1, \ldots, b_m$, таких, что $b_1 \leqslant b_2 \leqslant \ldots \leqslant b_m$. Будем говорить, что элементы $b_1, \ldots, b_m$ последовательно покрывают друг друга, если $b_1 \prec b_2 \prec \ldots \prec b_m$.

Теорема 2. Конечная решётка S изоморфна MLS-решётке, то есть прямому произведению полной решётки подмножеств на линейную решётку $SX(n) \times SL(m)$ для некоторых $m, n \geqslant 1$, тогда и только тогда, когда она удовлетворяет следующим трём условиям:

1) $|S| = 2^n m$;
2) если $m \geqslant 2$, то в точности $n + 1$ элементов решётки S являются атомарными; в случае $m = 1$ атомарны в точности n элементов S;
3) каждый элемент решётки S, кроме указанных далее $n+m$ элементов, покрывает по крайней мере два других. Исключения из этого условия составляют цепь из m последовательно покрывающих друг друга элементов S, начинающаяся с её инфинума O (каждый из них, кроме O, покрывает лишь предшествующий элемент цепи), а также n атомарных элементов, не входящих в эту цепь.

Доказательство. Проверим сначала выполнение условий 1–3 для решётки $T(n,m) = SX(n) \times SL(m)$. Условие 1 следует из того, что $|SX(n)| = 2^n$, $|SL(m)| = m$.

При $m = 1$ атомарными элементами решётки $T(n,1) = SX(n) \times SL(1) = SX(n)$ являются в точности 1-элементные подмножества множества $N = \{1, \ldots, n\}$, которых ровно n, что даёт выполнение условия 2.

Покажем, что при $m = 1$ все элементы решётки $SX(n)$, кроме атомарных и O (соответствующего пустому множеству), покрывают по крайней мере два других элемента $SX(n)$ (ранее показано, что атомарные элементы покрывают только O). Действительно, такие элементы являются k-элементными подмножествами множества N, $k \geqslant 2$. Из определения покрытия следует, что k-элементное множество покрывает все свои $(k-1)$-элементные подмножества. Число таких подмножеств равно $\mathrm{C}_k^{k-1} = k \geqslant 2$, что и требуется. Таким образом, при $m = 1$ условия 2 и 3 проверены.

Проверим условие 2 для решётки $T(n,m)$ при $m \geqslant 2$. Из условия (1) следует, что элемент $(a,b) \in T(n,m)$ является атомарным (то есть покрывает $(\varnothing, 0)$) тогда и только тогда, когда a атомарен в $SX(n)$ и $b = 0$, либо $a = \varnothing$ и b атомарен в $SL(m)$. Таким образом, атомарными в $T(n,m)$ являются в точности $n+1$ элементов $(\{1\}, 0), \ldots, (\{n\}, 0)$, $(\varnothing, 1)$, что доказывает выполнение условия 2.

Осталось проверить выполнение условия 3 при $m \geqslant 2$. Прежде всего заметим, что указанная в условии цепь из m последовательно покрывающих друг друга элементов получится, если рассмотреть элементы $(\varnothing, 0), \ldots, (\varnothing, m-1)$. Действительно, согласно условию (1), $(\varnothing, k) \succ (a,b)$ для некоторого $(a,b) \in T(n,m)$ равносильно тому, что $a = \varnothing$ (так как $a \prec \varnothing$ невозможно) и $b \prec k$, то есть $(a,b) = (\varnothing, k-1)$ — предыдущий элемент цепи. Таким образом, указанные элементы действительно образуют цепь и каждый элемент цепи (кроме $(\varnothing, 0)$, являющегося инфинумом $T(n,m)$) покрывает ровно один элемент $T(n,m)$ (предыдущий элемент цепи).

Покажем, что все элементы $T(n,m)$, кроме её инфинума, атомарных элементов и элементов рассмотренной выше цепи, покрывают по крайней мере два других элемента $T(n,m)$. Пусть $(a,b) \in T(n,m)$ — такой элемент ($a \subseteq N$, $b \in \{0, \ldots, m-1\}$). По построению это равносильно тому, что $|a| \geqslant 2$ при $b = 0$ и $a \neq \varnothing$ при $b \geqslant 1$. В первом случае, как мы видели при рассмотрении случая $m = 1$, найдутся такие $x, y \in SX(n)$, $x \neq y$, что $a \succ x$, $a \succ y$. Тогда $(a,0) \succ (x,0)$ и $(a,0) \succ (y,0)$. Во втором случае найдётся хотя бы один $x \in SX(n)$, такой, что $a \succ x$ (возможно, $x = \varnothing$), и тогда $(a,b) \succ (x,b)$ и $(a,b) \succ (a,b-1)$.

Докажем теперь обратное утверждение теоремы — если решётка S удовлетворяет условиям 1–3 при некоторых $m, n \geqslant 1$, то она изоморфна $SX(n) \times SL(m)$.

Обозначим через $A = \{a_0, \ldots, a_n\}$ множество атомарных элементов S ($A = \{a_1, \ldots, a_n\}$ в случае $m = 1$). Если $m = 1$, указанная в условии 3 цепь по построению состоит только из элемента $O = \inf S$. В случае $m \geqslant 2$ эта цепь включает в себя m элементов, из которых ровно один (следующий за O) является атомарным; без ограничения общности считаем, что это a_0. Обозначим эту цепь $B = \{b_0, \ldots, b_{m-1}\}$, $b_0 \prec b_1 \prec \ldots \prec b_{m-1}$ (при этом $b_0 = O$, $b_1 = a_0$). Обозначим $A^* = A \setminus \{a_0\}$ ($A^* = A$ при $m = 1$). Отметим, что, независимо от значения m, $|A^*| = n$.

Зададим отображение $\psi : S \to 2^{A^*} \times B$ следующим образом. Каждому элементу $s \in S$ сопоставим упорядоченную пару, состоящую из подмножества всех элементов A^*, меньших или равных s, а также максимального элемента цепи B, меньшего или равного s:

$$\psi(s) = \left(\{a \in A^* : a \leqslant s\},\ \max\{b \in B : b \leqslant s\}\right).$$

Докажем, что отображение ψ является изоморфизмом решёток.

Покажем сначала, что ψ изотонно. Пусть $p, p' \in S$, $p \leqslant p'$. Обозначим $\psi(p) = (P, b)$, $\psi(p') = (P', b')$. Если для некоторого $a \in A^*$ выполнено $a \in P$, то $a \leqslant p \leqslant p'$ и $a \in P'$. Отсюда $P \subseteq P'$. Далее, из $b \leqslant p$ следует $b \leqslant p'$, поэтому $b \leqslant b'$. Тогда $(P, b) \leqslant (P', b')$ в решётке $2^{A^*} \times B$.

Покажем, что ψ инъективно. Пусть $\psi(p) = \psi(p') = (L, b_j)$ для $p, p' \in S$, выведем отсюда, что $p = p'$. Доказательство будем вести индукцией по сумме $|L| + j$.

При $|L| + j = 0$ имеем $\psi(p) = \psi(p') = (\varnothing, O)$, откуда следует, что в S нет ни одного атомного элемента, меньше или равного p или p'. Но тогда $p = p' = O$, так как если, например, $p > O$, то по лемме 1 в S найдётся элемент, меньший или равный p и покрывающий O.

Пусть утверждение выполнено для всех $|L|+j < k$, $k > 0$. Покажем его выполнение при $|L| + j = k$. Предположим противное: $\psi(p) = \psi(p') = (L, b_j)$, но $p \neq p'$. Обозначим $q = \sup(L \cup \{b_j\})$ (поскольку $p \neq p'$, без ограничения общности можно полагать $q \neq p$). По построению получаем $q \leqslant p$ (супремум элементов из $L \cup \{b_j\}$ меньше или равен их верхней грани) и, с учётом неравенства, $q < p$. Тогда по лемме 1 в интервале $[q, p]$ найдётся такой элемент r, что $r \succ q$. Заметим, что, исходя из построения q, $\psi(q) \geqslant (L, b_j)$. Тогда ввиду доказанной изотонности ψ получаем $\psi(q) = (L, b_j)$ в силу неравенства $q \leqslant p$ и затем $\psi(r) = (L, b_j)$ в силу неравенств $q \leqslant r \leqslant p$.

Поскольку $L \neq \varnothing$ или $b_j \neq O$, имеем $q \neq O$, а значит, r не атомарен. Кроме того, $r \notin B$, так как каждый элемент B, отличный от O, покрывает предшествующий в цепи, поэтому q также лежит в B, и $\psi(r) \neq \psi(q)$, поскольку $\psi(r) = (\cdot\,, r)$, а $\psi(q) = (\cdot\,, q)$. Тогда по условию 3 теоремы r покрывает по крайней мере два разных элемента S, и найдётся такой $t \in S$, $t \neq q$, что $r \succ t$. При этом $\psi(t) \leqslant \psi(r) = (L, b_j)$, но $\psi(t) \neq (L, b_j)$, так как в противном случае получим $t \geqslant q$ (из тех же соображений, из которых $p \geqslant q$), и тогда $r > t > q$, что противоречит тому, что r покрывает q. Таким образом, получили, что $\psi(t) < (L, b_j)$. Обозначим $\psi(t) = (M, b_i) < (L, b_j)$.

Из последнего неравенства следует, что $|M| + i < k$. Тогда для любой верхней грани t' множества $M \cup \{b_i\}$, меньше или равной t, из $t' \leqslant t$ ввиду изотонности ψ получим $\psi(t') \leqslant (M, b_i)$. С другой стороны, для t', являющегося верхней гранью $M \cup \{b_i\}$, по построению $\psi(t') \geqslant (M, b_i)$. Отсюда $\psi(t') = (M, b_i) = \psi(t)$, и $t' = t$ в силу предположения индукции. Заметим теперь, что t является верхней гранью (M, b_i) по построению. Тогда $\sup(M, b_i) \leqslant t$ и из проведёных рассуждений следует, что $t = \sup(M, b_i)$.

Но q также является верхней гранью $M \cup \{b_i\}$, и тогда $q \geqslant \sup(M, b_i) = t$. Получаем $r > q > t$ ($t \neq q$ по построению), а это противоречит тому, что r покрывает t. Инъективность ψ доказана.

Сюръективность ψ следует из его инъективности (дающей $|\psi(S)| = |S|$) и условия 1 теоремы (дающего $|S| = |2^{A^*} \times B|$). Таким образом, ψ задает изоморфизм решёток S и $2^{A^*} \times B$, последняя из которых изоморфна $SX(n) \times SL(m)$. ■

Сделаем замечание о структуре рассматриваемых теоремой 2 решёток, которое понадобится в дальнейшем.

Следствие 1. Решётки $SX(n) \times SL(2)$ и $SX(n+1)$, $n \geqslant 1$, эквивалентны.

Доказательство. Применим к решётке $S = SX(n) \times SL(2)$ прямое утверждение теоремы 2. Согласно условию 1, $|S| = 2^n \cdot 2 = 2^{n+1}$. Согласно условию 2, S содержит ровно $n + 1$ атомарных элементов. Наконец, указанная в условии цепь из двух элементов, помимо O, содержит элемент, покрывающий его, то есть атомарный. Таким образом, согласно условию 3, исключениями из правила, согласно которому каждый

элемент S должен покрывать два других её элемента, являются в точности O и все атомарные элементы S.

Теперь, применяя к S обратное утверждение теоремы 2, получаем с учётом доказанного, что S изоморфна $SX(n+1) \times SL(1)$, то есть $SX(n+1)$. ■

Оценим сложность проверки указанных в теореме 2 условий при определении структуры данной конечной решётки S. Решётку будем полагать заданной её матрицей инцидентности, то есть матрицей, для каждой пары элементов решётки указывающей, какой из них больше (либо что они несравнимы).

Теорема 3. Для проверки изоморфизма конечной решётки S, заданной своей матрицей инцидентности, произведению решёток $SX(n) \times SL(m)$ при каких-либо (неизвестных заранее) $m, n \geqslant 1$ достаточно выполнить $O(|S|^2)$ операций.

Доказательство. Будем проверять выполнение условий теоремы 2 для каких-либо m, n. Для этого определим элементы решётки S, покрывающие ровно один её элемент (а также определим инфинум S), затем выделим среди них атомарные элементы (что даст возможное значение n) и определим, удовлетворяют ли оставшиеся m элементов указанной в условии 3 теоремы 2 структуре.

Обозначим множество элементов S, покрывающих ровно один другой её элемент, через P. Покажем, что для определения принадлежности заданного элемента $p \in S$ множеству P достаточно совершить два прохода по набору $L(p) = \{q \in S : q < p\}$ элементов S, строго меньших p (они получаются из строки, соответствующей p в матрице инцидентности S). При этом если таких элементов нет ($L(p) = \varnothing$), то $p = O$. Если такой элемент только один, то $L(p) = \{O\}$ и p атомарен, так как если бы нашёлся такой $q \in S$, что $p > q > O$, то $O, q \in L(p)$. Множество атомарных элементов S обозначим A.

Рассмотрим случай $|L(p)| \geqslant 2$. Заметим, что $p \in P$ тогда и только тогда, когда в множестве $L(p)$ содержится его супремум (который p и покрывает). Действительно, если $r = \sup L(p)$, $r \in L(p)$, то $r < p$, и если найдётся такой $s \in S$, что $r < s < p$, то по построению $s \in L(p)$, и тогда $s \leqslant r$ — противоречие, означающее, что $r \prec p$. При этом для любого другого $r' \in L(p)$, $r' \neq r$, получается $r' < r < p$, что делает невозможным $r' \prec p$.

Обратно, пусть в $L(p)$ содержится ровно один элемент r, покрываемый p. Тогда в $L(p)$ не найдётся элемента, большего r, так как если $s > r$, $s \in L(p)$, то $p > s > r$ и p не покрывает r. Предположим сначала, что r является верхней гранью $L(p)$. Тогда $r = \sup L(p)$ (как и требуется), так как в противном случае найдётся меньшая верхняя грань $r' = \sup L(p)$, но тогда из $r \in L(p)$ следует $r \leqslant r'$ — противоречие. Если же предположить, что r не является верхней гранью $L(p)$, придём к противоречию. Действительно, в этом случае в $L(p)$ найдётся элемент s, $s \neq r$, не меньший чем r. По замечанию к лемме 1 тогда получим, что в интервале $[s, p]$ найдётся такой элемент r', что $r' \prec p$. При этом из $r' > s$ следует $r' \neq r$, и приходим к противоречию с тем, что p покрывает лишь один элемент решётки S.

Итак, мы показали, что условие $p \in P$ равносильно тому, что множество $L(p)$ содержит свой супремум (достаточно проверить содержание им своей верхней грани, так как в предыдущем абзаце показано, что любая верхняя грань $L(p)$, лежащая в $L(p)$, является его супремумом). Построим алгоритм, проверяющий условие наличия верхней грани множества $L(p)$ в нём самом.

За первый проход по множеству $L(p)$ находится его максимум q (элемент множества $L(p)$, не меньший никакого другого элемента этого множества).

За второй проход проверяем, является ли найденный максимум q верхней гранью $L(p)$, сравнивая его с остальными элементами. Если q не является верхней гранью $L(p)$, то $L(p)$ не содержит своей верхней грани, так как если r — верхняя грань $L(p)$, $r \in L(p)$, $r \neq q$, то $r > q$, что противоречит максимальности q.

Итак, два прохода по множеству $L(p)$ для каждого элемента $p \in S$ позволяют определить принадлежность $p \in P$ для всех p и, таким образом, дают множество P.

Если все элементы P атомарны, то, согласно условию 2 теоремы 2, возможны лишь случаи $m = 1$ или 2 (второй из которых сводится к первому ввиду доказанной в следствии теоремы 2 изоморфности решёток $SX(n+1)$ и $SX(n) \times SL(2)$). В этом случае берём, согласно условию 2 теоремы 2, $n = |A| - 1$, и если $|S| = 2^n$, то, применяя теорему с $m = 1$, получаем, что S изоморфна $SX(n)$. Если же $|S| \neq 2^n$, то, согласно условию 1 теоремы 2 (n определяется условием 2 однозначно), S не может быть изоморфна никакому произведению полной решётки подмножеств на линейную.

Рассмотрим случай, когда не все элементы P атомарны. Обозначим $B^* = P \setminus A$. Предполагая изоморфность S нужному прямому произведению, из условия 2 теоремы 2 вновь получаем $n = |A| - 1$ и, согласно условию 3, $m = |B^*| + 2$ (к неатомарным элементам P добавляется один атомарный, входящий в цепь из условия 3, и O). Теперь можем проверить равенство из условия 1: $|S| = 2^n m$ — если оно не выполнено, то требуемого изоморфизма нет.

Если это равенство выполнено, осталось проверить, что элементы B^* образуют указанную в условии 3 цепь, последовательно покрывая друг друга. Для этого на этапе построения записываем элементы B^* в массив, индексированный количеством элементов S, меньших заданного элемента B^*. Покажем, что если элементы B^* образуют требуемую цепь, эти индексы не могут совпадать друг с другом и образуют полный набор целых чисел от 2 до m.

Действительно, если $B^* = \{b_2, \ldots, b_{m-1}\}$, $O \prec b_1 \prec b_2 \prec b_3 \prec \ldots \prec b_{m-1}$, где b_1 — атомарный элемент, входящий в цепь, то каждый b_i превосходит i элементов $O, b_1, b_2, \ldots, b_{i-1}$. Для того чтобы показать, что b_i не может превосходить никаких других элементов S, предположим противное и выберем минимальный b_i, превосходящий какой-то элемент, кроме указанных. Обозначим такой элемент q. Тогда по замечанию к лемме 1 найдётся элемент $r \in [q, b_i]$, такой, что $r \prec b_i$. При этом $r \neq b_{i-1}$, так как в противном случае $q \leqslant r < b_{i-1}$, что противоречит минимальности выбранного b_i. Но тогда получаем, что b_i покрывает два разных элемента из S (r и b_{i-1}) — противоречие с условием 3 теоремы 2.

Итак, если указанные условия на индексы элементов B^* не выполнены, то S не изоморфна требуемому прямому произведению. Если условия выполнены, то для проверки условия 3 теоремы 2 и доказательства изоморфизма S решётке $SX(n) \times SL(m)$ с вычисленными выше n, m остаётся лишь проверить, что каждый следующий элемент массива больше предыдущего (что делается прямыми сравнениями за один проход по массиву). Действительно, в этом случае элементы B^* образуют цепь и каждый следующий элемент покрывает предыдущий, так как если $p > q$, но p не покрывает q, то есть $p > r > q$ для некоторого $r \in S$, то $L(p) \supseteq L(q) \cup \{q, r\}$ и $|L(p)| \geqslant |L(q)| + 2$, но по условиям на индексы $|L(p)| = |L(q)| + 1$.

Отдельно нужно рассмотреть минимальный элемент $b_2 \in B^*$ (с индексом 2 и соответствующей мощностью $|L(b_2)|$), который должен покрывать не другой элемент B^*, а элемент, не входящий в рассмотренный массив. Для него $L(b_2) = \{O, b_1\}$ при некотором $b_1 \in S$. Множество $L(b_2)$ имеет верхнюю грань $b_1 > O$, и тогда, как доказано выше, $b_2 \succ b_1$. При этом $b_1 \succ O$, так как в противном случае найдётся такой $r \in S$,

что $b_1 > r > O$, и тогда $O, b_1, r \in L(b_2)$, что противоречит условию $|L(b_2)| = 2$. Таким образом, недостающий фрагмент $O \prec b_1 \prec b_2$ цепочки $O \prec b_1 \prec \ldots \prec b_{m-1}$ построен.

Как видно из доказательства, для проверки изоморфизма требуется совершить два прохода по матрице инцидентности решётки S и некоторое количество существенно меньших по трудоёмкости операций. ■

3. Алгоритмы восстановления решётки

В терминологии теории графов определение 3 формулируется следующим образом: вершина v_1 решёточного графа G покрывает вершину v_2 этого же орграфа, если в транзитивном замыкании графа G существует дуга (v_1, v_2) и не существует ориентированного пути $p(v_1, v_2)$, такого, что $|p(v_1, v_2)| > 1$. Атомарная вершина (определение 4) имеет одну и только одну исходящую дугу, причём эта дуга должна вести в сток (вершину без исходящих дуг) орграфа. Будем говорить, что решёточный граф соответствует решётке S, если он порождает решётку, изоморфную решётке S. Переформулируем теорему 2 в терминах теории графов. При этом заметим, что так как исследуемый орграф решёточный, то сток t в нём существует и единственен.

Теорема 4. Решёточный граф соответствует MLS-решётке $S = SX(n_1) \times SL(n_2)$ тогда и только тогда, когда:

1) число вершин орграфа $n = 2^{n_1} n_2$;
2) все вершины орграфа покрывают по крайней мере две других, кроме вершин $x_1, \ldots, x_{n_1}$, t, $t_1, \ldots, t_{n_2-1}$. Если $n_2 = 1$, вершины t_j $(j = 1, \ldots, n_2 - 1)$ отсутствуют. При этом:
   - t — сток орграфа;
   - вершины $x_1, \ldots, x_{n_1}, t_1, \ldots, t_{n_2-1}$ покрывают ровно одну вершину;
   - вершины $x_1, \ldots, x_{n_1}, t_1$ атомарные;
   - вершины $t_{n_2-1} \succ \ldots \succ t_1 \succ t$ образуют цепь последовательно покрывающих друг друга вершин.

Для проверки соответствия решёточного графа решётке подмножеств теорема 4 может быть переформулирована следующим образом.

Теорема 5. Решёточный граф соответствует решётке подмножеств $SX(n_1)$ тогда и только тогда, когда он удовлетворяет следующим трём условиям:

1) число вершин орграфа $n = 2^{n_1}$;
2) число атомарных вершин равно $\log_2 n = n_1$;
3) все вершины, кроме атомарных и стока, покрывают не менее двух других.

Представим алгоритм 1 проверки указанных в теореме 4 условий для определения структуры заданного решёточного графа.

Пусть решёточный граф G задан матрицей достижимости $\mathbf{M}^+$ размерности $n \times n$. Требуется получить ответ на вопрос, соответствует ли орграф G некоторой MLS-решётке $S = SX(n_1) \times SL(n_2)$. Для этого необходимо найти вершины орграфа G, покрывающие ровно одну вершину. Пусть эти вершины образуют множество P, $|P| = p$. Среди вершин множества P необходимо найти атомарные. Обозначим множество этих вершин A, $|A| = a$. Очевидно, что должны выполняться следующие равенства:

$$p = n_1 + n_2 - 1, \quad a = \begin{cases} n_1, & n_2 = 1, \\ n_1 + 1, & n_2 > 1. \end{cases}$$

Пусть $C = P \setminus A$ — множество вершин, покрывающих ровно один элемент, но не являющихся атомарными.

Алгоритм 1. Определение структуры решёточного графа

Вход: решёточный граф G задан матрицей достижимости $\mathbf{M}^+$ размерности $n \times n$; вспомогательный массив $\mathbf{V}$ размерности n.

Выход: соответствует ли орграф G некоторой MLS-решётке $S = SX(n_1) \times SL(n_2)$?

1: Сформируем множества P, A и C. Для этого
2: **Для всех** вершин v_i орграфа G
3: По строке i матрицы достижимости совершим проход и найдём максимальный элемент среди тех, которые формируют множество $L(v_i)$; обозначим его v_* (напомним, что $v_j \geqslant v_k \Leftrightarrow [\mathbf{M}^+]_{jk} = 1$).
4: Ещё за один проход по i-й строке матрицы проверим, что v_* является верхней гранью для $L(v_i)$ ($[\mathbf{M}^+]_{*j} = 1$ для всех $v_j \in L(v_i)$).
5: **Если** в i-й строке матрицы единичный элемент один (он же является верхней гранью), **то**
6: $v_i \in P$ и $v_i \in A$.
7: **Если** единичных элементов более одного и верхняя грань найдена, **то**
8: $v_i \in P$ и $v_i \in C$. Пополним элемент массива $[\mathbf{V}]_q$ вершиной v_i, где индекс q — это мощность множества $L(v_i)$.
9: Проанализируем множества P и A. Возможны два случая.
10: **Если** $A = P$, **то**
все вершины множества P атомарны. Тогда $n_1 = p$, $n_2 = 1$.
11: **Если** $n \neq 2^p$ (условие 1 теоремы 4), **то**
12: требуемого соответствия нет,
13: **иначе**
14: орграф G соответствует решётке подмножеств $S = XS(p)$ (или эквивалентной MLS-решётке $S = XS(p-1) \times LS(2)$ (следствие 1)).
15: **Если** $A \subset P$, **то**
16: не все вершины множества P атомарны. Тогда $n_1 = a - 1$, $n_2 = p - a + 2$.
17: **Если** $n \neq 2^{a-1}(p - a + 2)$ (условие 1 теоремы 4), **то**
18: требуемого соответствия нет,
19: **иначе**
20: **Если** все элементы массива $\mathbf{V}$ в диапазоне индексов от 2 до $n_2 - 1$ содержат ровно одну вершину, каждая следующая вершина этого массива в заданном диапазоне покрывает предыдущую, вершина $[\mathbf{V}]_2$ покрывает некоторую атомарную вершину из множества A, **то**
21: орграф G соответствует MLS-решётке $S = XS(a-1) \times LS(p-a+2)$,
22: **иначе**
23: требуемого соответствия нет.

Прокомментируем шаги 2–10 алгоритма 1. В них $L(v_i) = \{v_j : v_i > v_j\}$ — множество вершин, до которых существует ориентированный путь из вершины v_i. Для этих вершин $[\mathbf{M}^+]_{ij} = 1$, то есть множество $L(v_i)$ определяется единичными недиагональными элементами строки i матрицы достижимости. При доказательстве теоремы 3 показано, что для проверки принадлежности вершины v_i множеству P достаточно выяснить, содержит ли множество $L(v_i)$ свою верхнюю грань. Если $|L(v_i)| = 1$, то v_i — атомарная вершина.

После того как множества P, A и C сформированы, необходимо проанализировать их структуру (шаги 11–25 алгоритма 1). Пояснений требует случай, когда $A \subset P$ и

$n = 2^{a-1}(p - a + 2)$ (шаг 22). В этой ситуации остаётся проверить, что все вершины из множества C образуют указанную в условии 2 теоремы 4 цепь, последовательно покрывая друг друга. Это означает, что число вершин, в которые существует ориентированный путь из вершины v_i, для всех вершин множества C уникально. Более того, эти величины образуют полный набор чисел в диапазоне от 2 до $n_2 - 1$ (см. доказательство теоремы 3). Для проверки этого факта на этапе построения множества C каждую его вершину v_i необходимо продублировать в массив, индексированный числом вершин, в которые существует ориентированный путь из v_i (это число единиц в строке i матрицы достижимости, или мощность множества $L(v_i)$). Индексы k этого массива меняются от 2 до $n_2 - 1$. Каждый элемент массива должен содержать ровно одну вершину. Если указанное условие не выполнено, то искомого соответствия нет. Иначе осталось проверить, что каждая следующая вершина массива покрывает предыдущую: $v_{i_{k+1}} \succ v_{i_k}$. Для этого достаточно, чтобы $[\mathbf{M}^+]_{i_{k+1}i_k} = 1$. Отдельно необходимо рассмотреть вершину v_{i_2}, которая должна покрывать не другую вершину множества C, а некоторую атомарную вершину из множества A (что также проверяется по матрице достижимости: $[\mathbf{M}^+]_{i_2 j} = 1$ для некоторой вершины $v_j \in A$).

Замечание 1. Алгоритм 1 очевидным образом можно адаптировать для проверки соответствия решёточного графа решётке подмножеств.

Замечание 2. Трудоёмкость проверки соответствия решёточного графа, заданного матрицей смежности, некоторой MLS-решётке $S = XS(n_1) \times LS(n_2)$ при неизвестных заранее n_1 и n_2 в общем случае не превосходит $\mathrm{O}(n^3)$, где n — число вершин орграфа. Действительно, в теореме 3 доказано, что сложность проверки не превосходит $\mathrm{O}(n^2)$ для матрицы инцидентности исследуемой решётки. При теоретико-графовой интерпретации эта матрица совпадает с матрицей достижимости $\mathbf{M}^+$ орграфа. Если исследуемый решёточный граф задан матрицей смежности, то трудоёмкость повышается до $\mathrm{O}(n^3)$, так как «узким» местом проверки будет алгоритм Уоршелла, позволяющий построить матрицу достижимости.

Следует заметить, что при оценке трудоёмкости алгоритмов операции по формированию множеств считались линейными, тогда как трудоёмкость подобных операций существенно зависит от реализации структур данных.

Теоремы 2 и 3 позволили построить алгоритм проверки соответствия решёточного графа MLS-решётке (в частном случае, решётке подмножеств). Исследуем случай линейной решётки.

Теорема 6. Трудоёмкость проверки соответствия решёточного графа некоторой линейной решётке не превосходит $\mathrm{O}(n^3)$.

Доказательство. Для проверки соответствия решёточного графа линейной решётке достаточно показать, что любая пара вершин сравнима, то есть для любой пары вершин существует ориентированный путь, их соединяющий. Очевидно, трудоёмкость алгоритма определяется алгоритмом Уоршелла построения матрицы достижимости. ■

Заключение

Доказано, что задача восстановления решётки ценностей по орграфу разрешима за полиномиальное время. Этот факт существенен при построении систем разграничения доступа, основанных на мандатной модели безопасности [8], так как гарантирует получение результата за приемлемое время.

Авторы выражают благодарность Ю. С. Ракицкому за полезные замечания и предложения в процессе подготовки статьи.

ЛИТЕРАТУРА

1. *Belim S., Bogachenko N., and Ilushechkin E.* An analysis of graphs that represent a role-based security policy hierarchy // J. Computer Security. 2015. V. 23. No. 5. P. 641–657.
2. *Birkhoff G.* Lattice Theory. N.Y.: Amer. Math. Soc. Colloquium Publ., 1967.
3. *Jakubik J.* On isomorphisms of graphs of lattices // Czechoslovak Math. J. 1985. V. 35. No. 2. P. 188–200.
4. *Mainetti M.* Arguesian identities in linear lattices // Adv. Math. 1999. No. 144. P. 50–93.
5. *Mainetti M. and Yan C. H.* Geometric identities in lattice theory // J. Combinatorial Theory. Ser. A. 2000. No. 91. P. 411–450.
6. *Felsner S. and Knauer K. B.* Distributive lattices from graphs // Proc. VI Jornadas de Matematica Discreta y Algoritmica. Lleida, 2008. P. 11–23.
7. *Bertet K.* The dependence graph of a lattice // Proc. CLA, Malaga, Spain, 2012. P. 223–231.
8. *Belim S. V., Bogachenko N. F., Kabanov A. N., and Rakitskiy Yu. S.* Using the Decision Support Algorithms Combining Different Security Policies // Dynamics of Systems, Mechanisms and Machines (Dynamics). 15–17 Nov. 2016. IEEE Xplore. `http://ieeexplore.ieee.org/document/7818976/`.

REFERENCES

1. *Belim S., Bogachenko N., and Ilushechkin E.* An analysis of graphs that represent a role-based security policy hierarchy. J. Computer Security, 2015, vol. 23, no. 5, pp. 641–657.
2. *Birkhoff G.* Lattice Theory. N.Y., Amer. Math. Soc. Colloquium Publ., 1967.
3. *Jakubik J.* On isomorphisms of graphs of lattices. Czechoslovak Math. J., 1985, vol. 35, no. 2, pp. 188–200.
4. *Mainetti M.* Arguesian identities in linear lattices. Adv. Math., 1999, no. 144, pp. 50–93.
5. *Mainetti M. and Yan C. H.* Geometric identities in lattice theory. J. Combinatorial Theory, Ser. A, 2000, no. 91, pp. 411–450.
6. *Felsner S. and Knauer K. B.* Distributive lattices from graphs. Proc. VI Jornadas de Matematica Discreta y Algoritmica, Lleida, 2008, pp. 11–23.
7. *Bertet K.* The dependence graph of a lattice. Proc. CLA, Malaga, Spain, 2012, pp. 223–231.
8. *Belim S. V., Bogachenko N. F., Kabanov A. N., and Rakitskiy Yu. S.* Using the Decision Support Algorithms Combining Different Security Policies. Dynamics of Systems, Mechanisms and Machines (Dynamics), 15–17 Nov. 2016, IEEE Xplore. `http://ieeexplore.ieee.org/document/7818976/`.

УДК 519.1

СТРУКТУРНЫЕ СВОЙСТВА МИНИМАЛЬНЫХ ПРИМИТИВНЫХ ОРГРАФОВ

Ф. В. Лебедев

ООО «АСП Лабс», г. Москва, Россия

Описаны классы n-вершинных минимальных примитивных орграфов с числом дуг $(n+3)$, приведены их степенные структуры. Установлена зависимость структурных свойств n-вершинных минимальных примитивных орграфов от числа дуг. В частности, получена оценка количества классов таких графов с $(n+k)$ дугами.

Ключевые слова: *примитивная матрица, примитивный орграф, сильносвязный орграф.*

DOI 10.17223/20710410/41/7

STRUCTURAL PROPERTIES OF MINIMAL PRIMITIVE DIGRAPHS

P. V. Lebedev

Ltd «ASP Labs», Moscow, Russia

E-mail: plebedev@asplabs.ru

Let $\Gamma^P(n,m)$ be the set of all minimal primitive n-vertex digraphs with m arcs. The purpose of the research is to describe the new classes of digraphs $\Gamma \in \Gamma^P(n,n+3)$ and their graph degree structures $D(\Gamma)$. This problem is important for the analysis of mixing properties of round transformations, e.g. symmetric iterative block ciphers. A matrix M is said to be primitive if there is a power $M^e = \left(m_{i,j}^{(e)}\right)$ such that $m_{i,j}^{(e)} > 0$ for all i and j; the least power e with this property is called an exponent of M. The conceptions of the primitiveness and exponent of the matrix M expand to the digraph Γ with the adjacency matrix M. The minimal primitive digraph is a digraph of which adjacency matrix loses its primitiveness property after replacing any positive element by zero. The main results of our research are the following: 1) for the minimal primitive digraph $\Gamma \in \Gamma^P(n,n+3)$, graph degree structures $D(\Gamma)$ are described via solutions of the equation $n_{1,2}+n_{2,1}+2n_{1,3}+2n_{2,2}+2n_{3,1}+3n_{1,4}+3n_{2,3}+ +3n_{3,2}+3n_{4,1}+\ldots+(n-2)n_{n-1,1}=6$ and represented in the table of $D(\Gamma)$ values; 2) it is proved that $D(\Gamma)$, for digraphs from the set $\Gamma^P(n,n+k)$, are determined and can be calculated by $D(\Gamma)$ for $\Gamma \in \Gamma^P(n-1,n+k-2)$; 3) it is proved that the number of classes of digraphs $\Gamma^P(n,n+k)$ could be estimated via solutions of the equation $n_{1,2}+n_{2,1}+2n_{1,3}+2n_{3,1}+3n_{1,4}+3n_{4,1}+4n_{1,5}+4n_{5,1}+\ldots+kn_{1,k+1}+kn_{k+1,1}=2k$ and graph degree structures for $\Gamma \in \Gamma^P(n-1,n+k-2)$; 4) $N_3 \leqslant 34$ and $N_2 \leqslant 9$, where N_i is the number of classes in $\Gamma^P(n,n+i)$.

Keywords: *primitive matrix, primitive digraph, strongly connected digraph.*

Введение

Зачастую криптографические преобразования представляют собой систему булевых функций, заданную координатными функциями $f_1(x_1, \ldots, x_m), \ldots, f_n(x_1, \ldots, x_m)$. Особенность такой системы в том, что её надёжность напрямую зависит от перемешивания входов. Чем больше существенных переменных у каждой координатной функции системы, тем она надёжнее. Наилучший эффект достигается тогда, когда каждая координатная функция существенно зависит от каждой переменной, в таком случае имеется так называемое полное перемешивание входов. Перемешивание входов можно охарактеризовать с помощью ориентированного графа, которому можно сопоставить неотрицательную матрицу, называемую матрицей смежности графа, поэтому для исследования связей между элементами удобно применять матрично-графовый подход [1]. Обзор известных результатов по этому направлению дан в [2]. Сложность реализации системы характеризуется, в частности, числом связей (дуг орграфа Γ). Минимальные примитивные матрицы (МПМ) и орграфы (МПО) представляют интерес с точки зрения экономной реализации коммуникативной системы. Результаты исследования МПО содержатся в [3], где описаны структурные свойства n-вершинных МПО с $(n+1)$ и $(n+2)$ дугами.

В данной работе проведено исследование структурных свойств n-вершинных МПО с $(n+3)$ дугами, что является расширением известных результатов и логическим продолжением [3]. Основные обозначения, используемые в работе:

- $\Gamma^P(n, m)$ — множество минимальных примитивных орграфов с числом вершин n и числом дуг m;
- K^* — система контуров;
- $D(\Gamma)$ — степенная структура орграфа Γ;
- $n_{r,s}$ — количество вершин с полустепенью захода r и полустепенью исхода s;
- p_i — полустепень захода вершины i;
- q_i — полустепень исхода вершины i;
- $[i, j]$ — простой путь в орграфе Γ из вершины i в вершину j.

1. Подход к описанию структурных свойств минимальных примитивных орграфов

Заметим, что матрица и соответствующий ей граф одновременно либо примитивны, либо непримитивны, поэтому в работе используется язык теории графов. В [3] вводится понятие степенной структуры орграфа — таблицы положительных чисел $n_{r,s}$ при всех допустимых значениях r и s, описывающих количество заходящих и исходящих дуг вершины n. В [3] также впервые вводятся понятия примитивной (минимальной примитивной) системы контуров — такой, что натянутый на неё подграф примитивен, и K^*-изолированной дуги — не принадлежащей ни одному из контуров системы K^*. В [3] доказаны теоремы, являющиеся основными в области изучения структурных свойств МПО.

Теорема 1 [3]. Если граф $\Gamma \in \Gamma^P(n, m)$ при некоторых натуральных n и m, K^* — примитивная (минимальная примитивная) система контуров в Γ и в Γ имеется K^*-изолированная дуга, то при любом натуральном k имеется орграф Γ_k из $\Gamma^p(n+k, m+k)$, являющийся k-расширением графа Γ и содержащий систему K^*. Если при этом орграф Γ минимальный, то имеется k-расширение Γ_k, являющееся минимальным примитивным графом.

Теорема 2 [3]. При $n \geqslant 3$ орграф $\Gamma \in \Gamma^p(n, n+1)$ тогда и только тогда, когда Γ есть объединение простых контуров взаимно простых длин l и λ, общая часть которых есть путь длины q, где $l > \lambda$; $l + \lambda - q = n + 1$; $0 \leqslant q \leqslant n - 2$; при $q = 2$ общая часть контуров есть вершина.

Классы n-вершинных МПО с $(n+k)$ дугами можно описать системой из двух уравнений, одно из которых перечисляет удвоенное число дуг в графе (в соответствии с теоремой Эйлера [4]), а другое — число вершин в данном графе:

$$2n_{1,1} + 3n_{1,2} + 3n_{2,1} + 4n_{1,3} + 4n_{2,2} + 4n_{31} + \ldots + nn_{n-1,1} = 2(n + k); \quad (1)$$

$$n_{1,1} + n_{1,2} + n_{2,1} + n_{1,3} + n_{2,2} + n_{3,1} + n_{1,4} + n_{2,3} + n_{3,2} + n_{4,1} + \ldots + n_{n-1,1} = n. \quad (2)$$

Систему, состоящую из уравнений (1) и (2), обозначим (*). Решив её, можно описать все классы минимальных примитивных орграфов, принадлежащие $\Gamma^P(n, n+k)$. Заметим, что уравнения системы (*) относятся к классу диофантовых уравнений, описанных в [5], однако интерес представляют только неотрицательные решения, поскольку они являются количественными характеристиками графа. Вычитая из уравнения (1) удвоенное уравнение (2), получим

$$n_{1,1} + n_{2,1} + 2n_{1,3} + 2n_{2,2} + 2n_{3,1} + 3n_{1,4} + 3n_{2,3} + 3n_{3,2} + 3n_{4,1} + \ldots + (n-2)n_{n-1,1} = 2k. \quad (3)$$

При $k = 1$ уравнение (3) имеет вид

$$n_{1,2} + n_{2,1} + 2n_{1,3} + 2n_{2,2} + 2n_{3,1} = 2. \quad (4)$$

Уравнение (4) описывает случай, когда число дуг превосходит число вершин орграфа на единицу и имеет два решения в целых неотрицательных числах, соответственно имеется два класса $\Gamma^P(n, n+1)$.

1 к л а с с: $n_{1,2} = n_{2,1} = n_{1,3} = n_{3,1} = 0$, $n_{2,2} = 1$. Пример графа приведен на рис. 1.

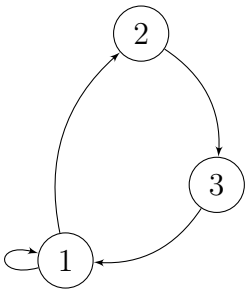

Рис. 1. Граф Γ, $n = 3$, $D(\Gamma) = \{(1,1)^2, (2,2)^1\}$

В соответствии с теоремой 1 степенная структура графов первого класса имеет следующий вид: $D(\Gamma_k) = \{(1,1)^{k+2}, (2,2)^1\}$.

2 к л а с с: $n_{1,2} = n_{2,1} = 1, n_{1,3} = n_{3,1} = 0$. Пример графа представлен на рис. 2. Степенная структура графов второго класса имеет вид $D(\Gamma_k) = \{(1,1)^{k+2}, (1,2)^1, (2,1)^1\}$.

Применяя данный подход к исследованию структурных свойств МПО, приведём ещё одну формулировку теоремы 2.

Теорема 3. Если минимальный примитивный орграф $\Gamma \in \Gamma^P(n, n+1)$, то $D(\Gamma)$ принадлежит классам, описанным в табл. 1.

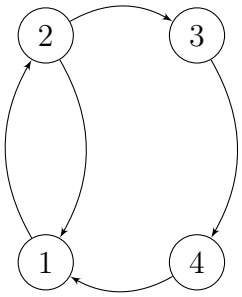

Рис. 2. Граф Γ, $n = 4$, $D(\Gamma) = \{(1,1)^2, (1,2)^1, (2,1)^1\}$

Таблица 1

Классы минимальных примитивных орграфов $\Gamma \in \Gamma^p(n, n+1)$

№ п/п	n	$D(\Gamma)$
1	$\geqslant 3$	$\{(1,1)^{n-1}, (2,2)^1\}$
2	$\geqslant 4$	$\{(1,1)^{n-2}, (1,2)^1, (2,1)^1\}$

Данный подход впервые предложен в [3] и отражён в теореме 4.

Теорема 4. Если минимальный примитивный орграф $\Gamma \in \Gamma^P(n, n+2)$, то $D(\Gamma)$ принадлежит классам, описанным в табл. 2.

Таблица 2

Классы минимальных примитивных орграфов $\Gamma \in \Gamma^p(n, n+2)$

№ п/п	n	$D(\Gamma)$
1	$\geqslant 5$	$\{(1,1)^{n-1}, (3,3)^1\}$
2	$\geqslant 5$	$\{(1,1)^{n-2}, (2,1)^1, (2,3)^1\}$
3	$\geqslant 5$	$\{(1,1)^{n-2}, (1,2)^1, (3,2)^1\}$
4	$\geqslant 5$	$\{(1,1)^{n-2}, (2,2)^2\}$
5	$\geqslant 4$	$\{(1,1)^{n-3}, (3,1)^1, (1,3)^1\}$
6	$\geqslant 6$	$\{(1,1)^{n-3}, (1,3)^1, (2,1)^2\}$
7	$\geqslant 6$	$\{(1,1)^{n-3}, (1,2)^2, (3,1)^1\}$
8	$\geqslant 6$	$\{(1,1)^{n-3}, (2,1)^1, (1,2)^1, (2,2)^1\}$
9	$\geqslant 6$	$\{(1,1)^{n-4}, (2,1)^2, (1,2)^2\}$

Заметим, что с ростом разницы числа дуг и вершин количество решений значительно увеличивается, соответственно сложность описания классов МПО возрастает.

Так как в любом графе сумма всех полустепеней исхода равна сумме всех полустепеней захода, можно сказать, что числа $n_{r,s}$ связаны следующим равенством:

$$\sum k_{r_i,s_i} r_i = \sum k_{r_i,s_i} s_i, \tag{5}$$

где k_{r_i,s_i} — коэффициент при n_{r_i,s_i}, $k_{r_i,s_i} = r_i + s_i - 2$.

2. Структурные свойства n-вершинных МПО с $(n+3)$ дугами

Теорема 5. Если минимальный примитивный орграф $\Gamma \in \Gamma^P(n, n+3)$, то $D(\Gamma)$ принадлежит классам, приведённым в табл. 3.

Таблица 3

Классы минимальных примитивных орграфов $\Gamma \in \Gamma^p(n, n+3)$

№ п/п	n	$D(\Gamma)$
1	$\geqslant 6$	$\{(1,1)^{n-1}, (4,4)^1\}$
2	$\geqslant 6$	$\{(1,1)^{n-2}, (2,1)^1, (3,4)^1\}$
3	$\geqslant 6$	$\{(1,1)^{n-2}, (1,2)^1, (4,3)^1\}$
4	$\geqslant 6$	$\{(1,1)^{n-2}, (2,2)^1, (3,3)^1\}$
5	$\geqslant 7$	$\{(1,1)^{n-3}, (2,1)^1, (1,2)^1, (3,3)^1\}$
6	$\geqslant 5$	$\{(1,1)^{n-2}, (1,3)^1, (4,2)^1\}$
7	$\geqslant 7$	$\{(1,1)^{n-3}, (1,2)^2, (4,2)^1\}$
8	$\geqslant 6$	$\{(1,1)^{n-2}, (3,1)^1, (2,4)^1\}$
9	$\geqslant 7$	$\{(1,1)^{n-3}, (2,1)^2, (2,4)^1\}$
10	$\geqslant 6$	$\{(1,1)^{n-3}, (2,2)^1, (1,2)^1, (3,2)^1\}$
11	$\geqslant 8$	$\{(1,1)^{n-4}, (2,1)^1, (1,2)^2, (3,2)^1\}$
12	$\geqslant 6$	$\{(1,1)^{n-2}, (2,3)^1, (3,2)^1\}$
13	$\geqslant 7$	$\{(1,1)^{n-3}, (1,3)^1, (2,1)^1, (3,2)^1\}$
14	$\geqslant 6$	$\{(1,1)^{n-3}, (2,2)^1, (2,1)^1, (2,3)^1\}$
15	$\geqslant 8$	$\{(1,1)^{n-4}, (1,2)^1, (2,1)^2, (2,3)^1\}$
16	$\geqslant 7$	$\{(1,1)^{n-3}, (3,1)^1, (1,2)^1, (2,3)^1\}$
17	$\geqslant 7$	$\{(1,1)^{n-4}, (2,1)^3, (1,4)^1\}$
18	$\geqslant 7$	$\{(1,1)^{n-3}, (2,1)^1, (3,1)^1, (1,4)^1\}$
19	$\geqslant 7$	$\{(1,1)^{n-4}, (1,2)^3, (4,1)^1\}$
20	$\geqslant 7$	$\{(1,1)^{n-3}, (1,2)^1, (1,3)^1, (4,1)^1\}$
21	$\geqslant 6$	$\{(1,1)^{n-3}, (2,2)^3\}$
22	$\geqslant 7$	$\{(1,1)^{n-4}, (2,2)^2, (1,2)^1, (2,1)^1\}$
23	$\geqslant 5$	$\{(1,1)^{n-4}, (2,2)^2, (1,3)^1, (3,1)^1\}$
24	$\geqslant 7$	$\{(1,1)^{n-5}, (2,2)^1, (1,2)^2, (2,1)^2\}$
25	$\geqslant 7$	$\{(1,1)^{n-4}, (2,2)^1, (1,3)^1, (2,1)^2\}$
26	$\geqslant 7$	$\{(1,1)^{n-4}, (2,2)^1, (3,1)^2, (1,2)^2\}$
27	$\geqslant 7$	$\{(1,1)^{n-5}, (3,1)^1, (2,1)^1, (1,2)^3\}$
28	$\geqslant 6$	$\{(1,1)^{n-4}, (1,3)^1, (3,1)^1, (2,1)^1, (1,2)^1\}$
29	$\geqslant 8$	$\{(1,1)^{n-5}, (1,3)^1, (1,2)^1, (2,1)^3\}$
30	$\geqslant 9$	$\{(1,1)^{n-6}, (1,2)^3, (2,1)^3\}$

Доказательство. Если сильносвязный орграф $\Gamma \in \Gamma^P(n, n+3)$, то числа $n_{r,s}$ связаны системой (*) при $k = 3$. Составим уравнение, описывающее данные классы МПО, аналогично уравнению (4):

$$n_{1,2} + n_{2,1} + 2n_{1,3} + 2n_{2,2} + 2n_{3,1} + 3n_{1,4} + 3n_{2,3} + 3n_{3,2} + 3n_{4,1} + \ldots + (n-2)n_{n-1,1} = 6. \quad (6)$$

Определим решения уравнения (6) относительно целых неотрицательных чисел $n_{r,s}$ и укажем примитивные графы без петель, соответствующие полученным решениям. Заметим, что $n_{r,s} = 0$ при $r + s > 8$, следовательно, уравнение (6) равносильно следующему упрощённому уравнению:

$$\begin{gathered} n_{1,2} + n_{2,1} + 2n_{1,3} + 2n_{2,2} + 2n_{3,1} + 3n_{1,4} + 3n_{2,3} + 3n_{3,2} + 3n_{4,1} + 4n_{1,5} + 4n_{2,4} + \\ + 4n_{3,3} + 4n_{4,2} + 4n_{5,1} + 5n_{1,6} + 5n_{2,5} + 5n_{3,4} + 5n_{4,3} + 5n_{5,2} + 5n_{6,1} + 6n_{1,7} + \\ + 6n_{2,6} + 6n_{3,5} + 6n_{4,4} + 6n_{5,3} + 6n_{6,2} + 6n_{7,1} = 6. \end{gathered} \quad (7)$$

Пусть $n_{r,s} = 1$, а все остальные переменные равны нулю, тогда в Γ имеется вершина i, где $p_i = r$, $q_i = s$, то есть имеются дуги $(i, a), (i, b), \ldots, (i, r)$, где $i, a, b, \ldots, r$ различны. Орграф Γ — сильносвязный, значит, в Γ имеются простые пути $[a, i], [b, i], \ldots, [r, i]$.

Так как $p_i = s$, эти пути сходятся в s путей, при этом в Γ имеется вершина $j \neq i$, где $p_j \neq s$. Тогда при некоторых r и s имеем противоречие. Таким образом описываются классы МПО в [4]. Имеется 30 классов решения уравнения (7).

1-й к л а с с. Положим $n_{4,4} = 1$. В этом случае Γ есть объединение четырёх контуров, пересечение множеств вершин которых состоит из единственной вершины, а любая другая вершина принадлежит только одному из контуров. Пример графа приведён на рис. 3.

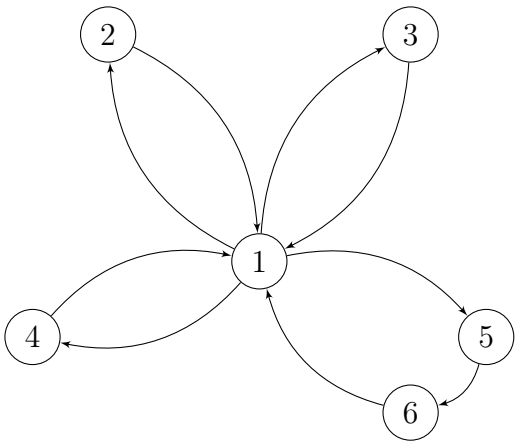

Рис. 3. Граф Γ, $n = 6$, $D(\Gamma) = \{(1,1)^5, (4,4)^1\}$

Степенная структура данного класса МПО имеет вид $D(\Gamma_k) = \{(1,1)^{k+5}, (4,4)^1\}$.

Если $n_{3,5} = 1$, то все остальные переменные в уравнении (7) равны нулю, следовательно, необходима ещё хотя бы одна вершина, чтобы уравнять количество входящих и исходящих дуг в графе. Отсюда следует, что если $n_{i,j} = 1$, $i + j = 8$, $i \neq j$, то уравнение (7) не имеет решений.

Рассмотрим упрощённое уравнение

$$\begin{gathered} n_{1,2} + n_{2,1} + 2n_{1,3} + 2n_{2,2} + 2n_{3,1} + 3n_{1,4} + 3n_{2,3} + 3n_{3,2} + 3n_{4,1} + 4n_{1,5} + 4n_{2,4} + \\ + 4n_{3,3} + 4n_{4,2} + 4n_{5,1} + 5n_{1,6} + 5n_{2,5} + 5n_{3,4} + 5n_{4,3} + 5n_{5,2} + 5n_{6,1} = 6, \end{gathered} \tag{8}$$

оно имеет ещё 29 решений. Путём аналогичных рассуждений получаем оставшиеся 29 классов. ■

3. Зависимость структурных свойств n-вершинных МПО от числа дуг

Определение 1. Назовём вершину, у которой полустепень исхода совпадает с полустепенью захода и равна 1, *моновершиной*.

Утверждение 1. Если существует решение уравнения (5) вида $n_{i_1,i_2} = a_1, \ldots, n_{i_m,i_m+1} = a_m$, то $n_{i_2,i_1} = a_1, \ldots, n_{i_m+1,i_m} = a_m$ также является решением уравнения (5).

Следствие 1. Графы, соответствующие таким решениям, изоморфны.

Утверждение 2. Класс МПО $\Gamma^P(n+p, n+p+k)$ образуется добавлением p вершин и $(p+1)$ дуг в соответствующие множества класса $\Gamma^P(n, n+k-1)$.

Доказательство. Количество вершин в полученном графе составит $(n+p)$, а количество дуг — $(n+k-1+p+1) = n+k+p$, отсюда следует, что количество дуг превышает количество вершин на k. ■

Пример 1. На рис. 4 изображён граф $\Gamma_1 \in \Gamma^P(5,7)$ со степенной структурой $D(\Gamma_1) = \{(1,1)^4, (3,3)^1\}$.

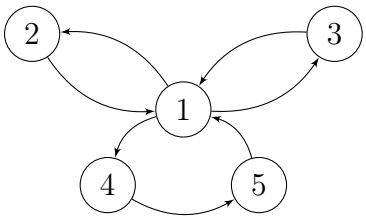

Рис. 4. Граф $\Gamma_1 \in \Gamma^P(5,7)$, $D(\Gamma_1) = \{(1,1)^4, (3,3)^1\}$

Добавим одну вершину и две дуги в соответствующие множества графа Γ_1 так, чтобы получить граф Γ_2, изображённый на рис. 5.

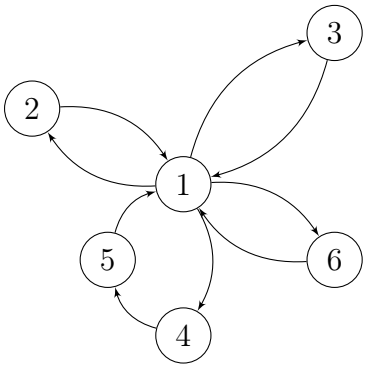

Рис. 5. Граф $\Gamma_2 \in \Gamma^P(6,9)$, $D(\Gamma) = \{(1,1)^5, (4,4)^1\}$

Как видно из рис. 4 и 5, графы Γ_1 и Γ_2 принадлежат множествам МПО, у которых разница между количеством дуг и вершин равна двум и трём соответственно, однако граф Γ_2 имеет на одну вершину больше. Заметим, что изменение степенной структуры говорит об увеличении количества вершин на 1 и количества дуг на 2.

Далее под *типом вершины* понимается характеристика вершины, описывающая количество заходящих и исходящих дуг данной вершины. Важно отметить, что в степенной структуре орграфа описываются все его типы вершин.

Пример 2. Рассмотрим степенную структуру орграфа Γ, принадлежащего 28-му классу $\Gamma^P(n, n+3)$, $D(\Gamma) = \{(1,1)^3, (1,3)^1, (3,1)^1, (2,1)^1, (1,2)^1\}$. Граф Γ имеет пять типов вершин: $(1,1), (1,3), (3,1), (2,1), (1,2)$.

Утверждение 3. Степенные структуры графов из множества $\Gamma^P(n, n+k)$ с точностью до количества моновершин определяются степенными структурами графов из множества $\Gamma^P(n-1, n+k-2)$.

Доказательство. Пусть a и b — вершины графа $\Gamma_1 \in \Gamma^p(n-1, n+k-2)$, при этом допустимо $a = b$. Так как добавляются две новые дуги (обозначим их A и B) и одна новая вершина (обозначим её d), то одна дуга является исходящей из этой вершины, а другая заходящей в неё, значит, $(p_d, q_d) = (1,1)$. Пусть дуга A исходит из a и заходит в d, а дуга B исходит из d и заходит в b, следовательно, степенная структура графа изменится и будет известна с точностью до количества моновершин в силу теоремы 1. ■

Утверждение 4. Степенная структура графа $\Gamma \in \Gamma^p(n, n+k)$ может быть получена из различных степенных структур графов множества $\Gamma^P(n-1, n+k-2)$.

Пример 3. Рассмотрим МПО $\Gamma_1, \Gamma_2 \in \Gamma^P(n, n+2)$, которые представлены на рис. 6, и их степенные структуры для $n = 5$: $D(\Gamma_1) = \{(1,1)^3, (2,1)^1, (2,3)^1\}$, $D(\Gamma_2) = \{(1,1)^3, (1,2)^1, (3,2)^1\}$.

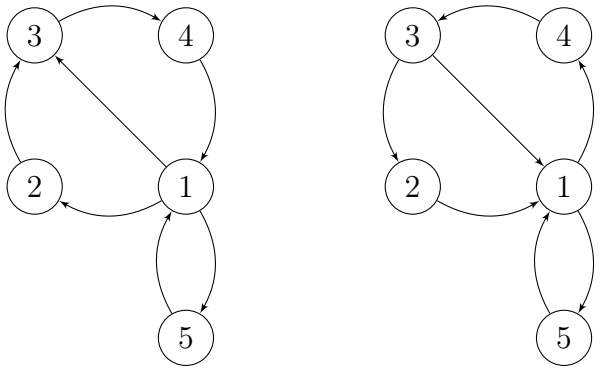

Рис. 6. $\Gamma_1, \Gamma_2 \in \Gamma^P(5, 7)$

Добавив к графу Γ_1 одну вершину и две дуги, можно получить граф Γ_3, такой, что $D(\Gamma_3) = \{(1,1)^3, (3,1)^1, (2,3)^1\}$. Также, определённым образом добавив к графу Γ_2 одну вершину и две дуги, можно получить граф Γ_4, такой, что $D(\Gamma_4) = \{(1,1)^3, (2,3)^1, (3,2)^1\}$. Заметим, что степенные структуры графов Γ_3 и Γ_4, которые представлены на рис. 7, совпадают.

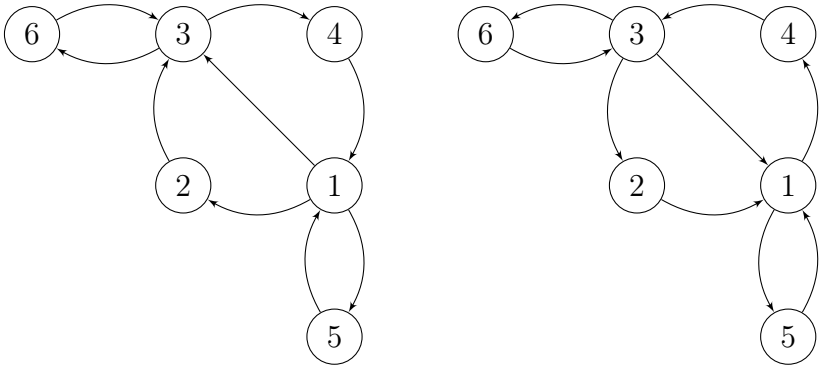

Рис. 7. $\Gamma_3, \Gamma_4 \in \Gamma^P(6, 9)$, $D(\Gamma_3) = D(\Gamma_4) = \{(1,1)^3, (2,3)^1, (3,2)^1\}$

Отметим, что графы Γ_3 и Γ_4 изоморфны и являются частными случаями графов 12-го класса n-вершинных МПО с количеством дуг $(n+3)$ для $n = 6$.

Утверждение 5. Количество классов $\Gamma^P(n, n+k)$ можно оценить с помощью степенных структур классов $\Gamma^P(n-1, n+k-2)$ и количества решений уравнения

$$n_{1,2} + n_{2,1} + 2n_{1,3} + 2n_{3,1} + 3n_{1,4} + 3n_{4,1} + 4n_{1,5} + 4n_{5,1} + \ldots + kn_{1,k+1} + kn_{k+1,1} = 2k. \quad (9)$$

Доказательство. Пусть $s_{k-1,i}$ — число типов вершин i-го класса из $\Gamma^P(n-1, n+k-2)$, тогда $\sum\limits_i s_{k-1,i}$ является оценкой сверху количества новых классов, не содержащих вершин, имеющих либо одну заходящую дугу, либо одну исходящую, за

исключением моновершин. С другой стороны, графы $\Gamma \in \Gamma^P(n, n+k)$ такого вида описываются уравнением (9). Пусть t_k — количество решений уравнения (9), N_k — количество классов $\Gamma^P(n, n+k)$. Зная степенные структуры классов $\Gamma^P(n, n+k-1)$, по утверждению 3 можно получить степенные структуры всех классов $\Gamma^P(n, n+k)$, при этом процесс подсчёта новых классов усложняется согласно утверждению 4. Отсюда следует, что $N_k \leqslant \sum\limits_i s_{k-1,i} + t_k$. ■

Пример 4. Рассмотрим классы $\Gamma^P(n, n+2)$. Согласно теореме 4, имеется 9 классов с различными степенными структурами и 26 типов вершин различных классов, описанных в табл. 2. Имеется 8 классов n-вершинных МПО с $(n+3)$ дугами, приведённых в табл. 3 и удовлетворяющих уравнению

$$n_{1,2} + n_{2,1} + 2n_{1,3} + 2n_{3,1} + 3n_{1,4} + 3n_{4,1} = n. \tag{10}$$

Данные классы описаны в табл. 4.

Таблица 4

Классы минимальных примитивных орграфов $\Gamma \in \Gamma^p(n, n+3)$, удовлетворяющих уравнению (10)

№ п/п	n	$D(\Gamma)$
1	$\geqslant 7$	$\{(1,1)^{n-4}, (2,1)^3, (1,4)^1\}$
2	$\geqslant 7$	$\{(1,1)^{n-3}, (2,1)^1, (3,1)^1, (1,4)^1\}$
3	$\geqslant 7$	$\{(1,1)^{n-4}, (1,2)^3, (4,1)^1\}$
4	$\geqslant 7$	$\{(1,1)^{n-3}, (1,2)^1, (1,3)^1, (4,1)^1\}$
5	$\geqslant 8$	$\{(1,1)^{n-5}, (3,1)^1, (2,1)^1, (1,2)^3\}$
6	$\geqslant 7$	$\{(1,1)^{n-4}, (1,3)^1, (3,1)^1, (2,1)^1, (1,2)^1\}$
7	$\geqslant 8$	$\{(1,1)^{n-5}, (1,3)^1, (1,2)^1, (2,1)^3\}$
8	$\geqslant 9$	$\{(1,1)^{n-6}, (1,2)^3, (2,1)^3\}$

Оценим количество классов $\Gamma^P(n, n+3)$, зная степенные структуры классов $\Gamma^P(n, n+2)$: по утверждению 5 верно $N_3 \leqslant 26+8 = 34$; и количество классов $\Gamma^P(n, n+2)$, зная степенные структуры классов $\Gamma^P(n, n+1)$: по утверждению 5 верно $N_2 \leqslant 5+4 = 9$. Заметим, что в данном случае полученная оценка полностью совпадает с количеством классов $\Gamma^P(n, n+2)$.

ЛИТЕРАТУРА

1. *Фомичев В. М.* Методы дискретной математики в криптологии. М.: Диалог-МИФИ, 2010. 424 с.
2. *Когос К. Г., Фомичев В. М.* Положительные свойства неотрицательных матриц // Прикладная дискретная математика. 2012. № 4 (18). С. 116–121.
3. *Фомичев В. М.* Свойства минимальных примитивных орграфов // Прикладная дискретная математика. 2015. № 2 (28). С. 86–96.
4. *Харари Ф.* Теория графов. М.: Едиториал УРСС, 2003. 296 с.
5. *Бухштаб А. А.* Теория чисел. СПб.: Лань, 2008. 384 с.

REFERENCES

1. *Fomichev V. M.* Metody diskretnoy matematiki v kriptologii [Methods of Discrete Mathematics in Cryptology]. Moscow, Dialog-MEPhI Publ., 2010. 324 p. (in Russian)

2. *Kogos K. G. and Fomichev V. M.* Polozhitel'nye svoystva neotritsatel'nykh matrits [Positive properties of non-negative matrices]. Prikladnaya Diskretnaya Matematika, 2012, no. 4(18), pp. 116–121. (in Russian)
3. *Fomichev V. M.* Svoystva minimal'nykh primitivnykh orgrafov [Properties of minimal primitive digraphs]. Prikladnaya Diskretnaya Matematika, 2015, no. 2(28), pp. 86–96. (in Russian)
4. *Harari F.* Graph Theory. Addison-Wesley, 1969.
5. *Bukhshtab A. A.* Teoriya chisel [Number Theory]. SPb., Lan' Publ., 2008. 384 p. (in Russian)

УДК 519.87

НОВЫЕ СЕМЕЙСТВА МУЛЬТИПЛИКАТИВНЫХ ЦИРКУЛЯНТНЫХ СЕТЕЙ[1]

Э. А. Монахова

Институт вычислительной математики и математической геофизики СО РАН, г. Новосибирск, Россия

Рассматривается задача оптимизации циркулянтных сетей, состоящая в максимизации числа вершин при заданных степени и диаметре графа. На основе изучения мультипликативных циркулянтов с образующими, представленными в виде степеней нечётных чисел $t \geqslant 5$, построены два новых семейства мультипликативных циркулянтов нечётных размерностей $k \geqslant 3$ и диаметров $d \equiv 0 \bmod k$ и чётных размерностей $k \geqslant 4$ и диаметров $d \equiv 0 \bmod k$ и $d \equiv 0 \bmod k/2$, графы которых превосходят по числу вершин при тех же размерностях и диаметрах известные семейства мультипликативных циркулянтов.

Ключевые слова: *мультипликативные циркулянтные сети, диаметр, максимальный порядок графа.*

DOI 10.17223/20710410/41/8

NEW FAMILIES OF MULTIPLICATIVE CIRCULANT NETWORKS

E. A. Monakhova

Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia

E-mail: emilia@rav.sscc.ru

For circulant networks, the problem of the maximal attainable number of nodes under given degree and diameter of their graphs is considered. A research of multiplicative circulant networks with generators in the form of $(1, t, t^2, \ldots, t^{k-1})$ for odd $t \geqslant 5$ is presented. On the base of this research, two new families of multiplicative circulant networks of orders $n = (t+1)(1+t+\ldots+t^{k-1})/2 + t^{k-1}$ for odd dimensions $k \geqslant 3$ and diameters $d \equiv 0 \bmod k$ and even dimensions $k \geqslant 4$ and diameters $d \equiv 0 \bmod k$ and $d \equiv 0 \bmod k/2$ are constructed. The orders of these graphs are larger than orders of graphs of all known families of multiplicative circulant networks under the same dimensions and diameters.

Keywords: *multiplicative circulant networks, diameter, maximum order of a graph.*

1. Введение и основные определения

Пусть $s_1, s_2, \ldots, s_k, n$ — целые числа, такие, что $1 \leqslant s_1 < s_2 < \ldots < s_k < n$. Граф C с множеством вершин $V = \{0, 1, \ldots, n-1\}$ и множеством рёбер $E = \{(i,j) : i - j \equiv \equiv s_l \bmod n,\ l = 1, \ldots, k\}$, называется *циркулянтным*, числа $S = (s_1, s_2, \ldots, s_k)$ — образующими, $(n; S)$ — его параметрическим описанием, k — размерностью, n — порядком

[1]Исследование выполнено в рамках проекта № 0315-2016-0006.

графа. Циркулянтные графы являются вершинно-транзитивными. Степень циркулянта равна $2k$, если $s_k \neq n/2$. При чётном n и $s_k = n/2$ циркулянт имеет степень $2k-1$. В данной работе рассматриваются только связные циркулянтные графы чётных степеней. Известно, что циркулянтный граф $C(n; s_1, s_2, \ldots, s_k)$ является связным, если и только если наибольший общий делитель чисел s_1, s_2, …, s_k, n равен 1.

Циркулянтные сети (графы) широко изучаются при проектировании и анализе вычислительных систем, в теории графов и дискретной математике, в качестве топологии мультипроцессорных систем и компьютерных сетей и для других применений [1–3]. Интересным представляется новое направление использования циркулянтных сетей в центрах обработки информации больших баз данных [4].

Циркулянтные сети $C(n; 1, t, t^2, \ldots, t^{k-1})$ с образующими, представленными в виде степеней натурального числа $t \geqslant 2$, называются *мультипликативными* циркулянтами. Следует отметить, что мультипликативные циркулянтные сети имеют простые коммуникационные алгоритмы, эффективны относительно трассировки интегральных схем, живучести и отказоустойчивости и поэтому могут использоваться в качестве структур сетей связи суперкомпьютерных систем.

Диаметром графа C называется $\mathrm{d}(C) = \max\limits_{i,j \in V} \mathrm{d}(i,j)$, где $\mathrm{d}(i,j)$ — длина кратчайшего пути из вершины i в вершину j графа C. Для любых натуральных d и k пусть $M(d,k)$ обозначает максимально возможное (достижимое) натуральное n, такое, что существует множество образующих $S = (1, s_2, \ldots, s_k)$, при котором $\mathrm{d}(C(n;S)) \leqslant d$. В [1, 2] можно найти обзоры результатов по оценкам диаметра и достижимого порядка k-мерных, $k \geqslant 2$, циркулянтных сетей.

Приведём известные результаты, касающиеся оценок диаметра и достижимого порядка k-мерных, $k \geqslant 2$, циркулянтных сетей.

В [5] показано, что $M(d,k) \leqslant 1 + \sum\limits_{i=0}^{k-1} \mathrm{C}_k^i \mathrm{C}_d^{k-i} 2^{k-i}$, получена нижняя граница диаметра для любых n и k порядка $\dfrac{1}{2}(k!)^{1/k} n^{1/k}$ и доказана

Теорема 1. Циркулянтные сети вида $C(n; 1, t, t^2, \ldots, t^{k-1})$, где $n = t^k$ и $t \geqslant 3$ — нечётное число, имеют диаметр $d = k\lfloor t/2 \rfloor$.

Для $k = 2$ задача построения семейств двумерных циркулянтов с единичной образующей и максимально возможным порядком при любом диаметре d решена (см., например, обзор в [1]). Найдена [6, 7] функция $M(d,k)$ для $k = 3$ и любого диаметра d и аналитически построены семейства трёхмерных циркулянтов с порядком, совпадающим с $M(d,3)$. Для $k = 4$ наилучшие известные оценки функции $M(d,4)$ аналитически найдены с помощью компьютерного поиска в [8]. В [9, 10] авторы представили таблицу, содержащую свод самых больших известных циркулянтных сетей, найденных в литературе для ряда значений степеней и диаметров. Таблица рекордных циркулянтов включает в том числе ряд значений порядков графов, полученных на основании перечисленных аналитических результатов для размерностей 2, 3 и 4, а также результаты для нечётных степеней, найденные в [7, 8, 10]. В настоящее время таблица рекордных циркулянтных сетей для $k \leqslant 8$ и $d \leqslant 10$ представлена в Интернете [11] и постоянно обновляется.

Изучение мультипликативных циркулянтов, начатое в работе [5], продолжалось в последующие годы. Результаты теоремы 1 улучшены в [12]:

Теорема 2. Пусть d и k — натуральные числа, $d \geqslant k \geqslant 3$ и $p = \lfloor (d-k+3)/k \rfloor$. Тогда

$$M(d,k) \geqslant n = 2p \sum_{i=0}^{k-1} (4p)^i = \frac{1}{2}\left(\frac{4}{k}\right)^k d^k + \mathrm{O}(d^{k-1}).$$

На основе теоремы 2 в [12] построены семейства мультипликативных циркулянтов с большим числом вершин, чем в [5], при тех же размерностях и диаметрах.

В [13, 14] рассмотрены свойства мультипликативных циркулянтных сетей вида $C(n; 1, t, \ldots, t^{k-1})$ с нечётным $t \geqslant 3$ и $2t^{k-1} < n \leqslant t^k$ и получена общая формула для верхней оценки диаметра:

$$d(n; 1, t, t^2, \ldots, t^{k-1}) \leqslant (k-1)\lfloor t/2 \rfloor + \lceil (n - t^{k-1})/(2t^{k-1}) \rceil.$$

В [15] исследованы свойства мультипликативных циркулянтов вида $C(n; 1, t, \ldots, t^{k-1})$ с $n = t^k$ и чётным $t \geqslant 2$ и получен их диаметр:

$$d(t^k; 1, t, t^2, \ldots, t^{k-1}) = kt/2 - \lfloor k/2 \rfloor.$$

Показано также, что мультипликативные циркулянтные сети как графы с образующими, представленными в виде степеней целого числа, имеют простые алгоритмы парного [13, 15] и трансляционного обменов [15], эффективны относительно трассировки интегральных схем, живучести и отказоустойчивости [13, 14].

В [16] получены новые улучшенные оценки для $M(d,k)$:

Теорема 3. Пусть $p = \left\lfloor \dfrac{d - \lfloor k/4 \rfloor}{k} \right\rfloor$, где d и $k > 4$ — целые числа, такие, что $d \geqslant k + \lfloor k/4 \rfloor$. Тогда

$$M(d,k) \geqslant n = \begin{cases} 2p \displaystyle\sum_{i=0}^{k-1} (4p)^i, & \text{если } kp + \lfloor k/4 \rfloor \leqslant d < kp + \lfloor k/2 \rfloor, \\ (2p+1) \displaystyle\sum_{i=0}^{k-1} (4p+1)^i, & \text{если } kp + \lfloor k/2 \rfloor \leqslant d < k(p+1), \\ (2p+2) \displaystyle\sum_{i=0}^{k-1} (4p+3)^i, & \text{если } k(p+1) \leqslant d < k(p+1) + \lfloor k/4 \rfloor. \end{cases}$$

В настоящей работе продолжено исследование нижних оценок экстремальной функции $M(d,k)$ и получение их аналитических выражений при любых размерностях $k > 4$. На основе изучения циркулянтов с образующими, представленными в виде степеней нечётных чисел $t \geqslant 5$, получены аналитически новые нижние оценки достижимого числа вершин циркулянтных сетей размерностей $k > 4$ и построены соответствующие семейства мультипликативных циркулянтов, реализующих эти оценки.

2. Новые семейства мультипликативных циркулянтных сетей

Для мультипликативных циркулянтов размерностей $k = 4$ и 5 проведены дальнейшие исследования сетей, рассмотренных в работах [5, 15, 16]. С помощью компьютерного поиска исследованы диапазоны их существования и определено максимальное значение порядка графа, при котором значения диаметров и образующих совпадают с найденными в [5, 15, 16]. Полученный результат позволил улучшить порядки графов по сравнению с известными результатами [5, 12, 15, 16]. Анализ значений подмножеств найденных максимально возможных порядков графов мультипликативных циркулянтов размерности 5 (табл. 1) послужил основой теоретического обобщения полученных результатов для любых размерностей.

В табл. 1 использованы следующие обозначения: d — диаметры найденных мультипликативных циркулянтов; n — их порядки; t — параметр, порождающий соответствующие множества образующих графа $S = (1, t, t^2, t^3, t^4)$.

Таблица 1

Новые циркулянтные графы размерности 5

$k = 5$								
d	n	t	d	n	t	d	n	t
6	682	4	15	111271	11	24	1245289	19
7	2343	5	16	137598	12	25	1505931	19
8	4399	7	17	216587	13	26	1692610	20
9	8803	7	18	282053	15	27	2246255	21
10	13605	7	19	383303	15	28	2671209	23
11	22820	8	20	484553	15	29	3230891	23
12	36905	9	21	563592	16	30	3790573	23
13	52707	11	22	798669	17	31	4168812	24
14	81989	11	23	984647	19	32	5289713	25

Рассмотрим множество мультипликативных циркулянтных сетей вида $C(n; 1, t, t^2, \ldots, t^{k-1})$, $k \geqslant 3$, с нечётным $t \geqslant 5$. В теореме 4 представлены два новых бесконечных семейства рассматриваемых сетей, которые улучшают известные оценки достижимого порядка циркулянтных графов. Далее $D(x)$, $0 \leqslant x < n$, обозначает длину кратчайшего пути из вершины 0 в вершину x.

Теорема 4. Пусть $k \geqslant 3$, $t \geqslant 5$ — нечётное число, $S = (1, t, t^2, \ldots, t^{k-1})$. Если

$$n = \lceil t/2 \rceil \sum_{i=0}^{k-1} t^i + t^{k-1}, \tag{1}$$

то

$$d(n; S) = \frac{k(t+1)}{4} \tag{2}$$

при следующих условиях:

$$k \geqslant 3 \text{ — нечётное число и } t \equiv 3 \bmod 4 \tag{3}$$

или

$$k \geqslant 4 \text{ — чётное число.} \tag{4}$$

Доказательство. Рассмотрим циркулянтный граф $C(n; 1, t, t^2, \ldots, t^{k-1})$, где $t \geqslant 5$ — нечётное число и значение n удовлетворяет (1). Заметим, что все вершины графа образуют замкнутый цикл $0, 1, \ldots, n-1, 0$.

Возьмём любую вершину $0 \leqslant x < n$ рассматриваемого графа. Определим c_0, такое, что $c_0 \equiv x \bmod t$ и $|c_0| \leqslant \lfloor t/2 \rfloor$. Для любых $i = 1, \ldots, k-1$ определим c_i, такие, что

$$c_i \equiv \frac{1}{t^i} \left(x - \sum_{j=0}^{i-1} c_j t^j \right) \bmod t,$$

$|c_i| \leqslant \lfloor t/2 \rfloor$ для $i = 0, \ldots, k-2$ и $0 \leqslant c_{k-1} \leqslant \lceil t/2 \rceil + 1$. Тогда

$$x = \sum_{i=0}^{k-1} c_i t^i = c_0 s_1 + c_1 s_2 + \ldots + c_{k-1} s_k.$$

Вычисленные таким образом коэффициенты c_i, $i = 0, \ldots, k-1$, являются координатами пути из вершины 0 в x, а именно: $|c_i|$ определяет, сколько раз в пути из вершины 0 в x использовалась соответствующая образующая, а знак (+ или −) у c_i указывает направление движения по соответствующей образующей. Обозначим длину этого пути через $D^+(x)$. Имеем $D^+(x) = \sum\limits_{i=0}^{k-1} |c_i|$.

Второй возможный путь из 0 в x определим, взяв разность $x - n$. Учитывая (1), имеем $x - n = \sum\limits_{i=0}^{k-1} c'_i t^i$, где $c'_i = c_i - \lceil t/2 \rceil$ для $i = 0, \ldots, k-2$ и $c'_{k-1} = c_{k-1} - \lceil t/2 \rceil - 1$.

В дальнейшем с помощью алгоритма 1 преобразуем коэффициенты c'_i в c''_i таким образом, чтобы выполнялось условие $|c''_i| \leqslant \lfloor t/2 \rfloor$, $i = 0, \ldots, k-2$. Преобразованные коэффициенты c''_i, $i = 0, \ldots, k-1$, являются координатами второго возможного пути из вершины 0 в x. Обозначим длину этого пути через $D^-(x)$. Она равна $D^-(x) = \sum\limits_{i=0}^{k-1} |c''_i|$.

Для любой вершины $0 \leqslant x < n$ длина кратчайшего пути

$$D(x) \leqslant \min\{D^+(x), D^-(x)\}.$$

Пусть значение выполнено (1). Покажем, что диаметр d графа $C(n; 1, t, t^2, \ldots, t^{k-1})$ удовлетворяет (2). Для этого докажем, что для любой вершины $0 \leqslant x < n$ при выполнении как условия (3), так и условия (4)

$$D^+(x) + D^-(x) \leqslant k\lceil t/2 \rceil + 1 = 2d + 1.$$

Далее последовательно для $i = 0, 1, \ldots, k-1$ выполняем преобразование коэффициентов c'_i в c''_i и подсчитываем суммы $|c_i| + |c''_i|$ (алгоритм 1).

На рис. 1 приведена граф-схема алгоритма 1, где $T = \lceil t/2 \rceil$.

Суммируя результаты выполнения алгоритма 1 и анализируя все возможности k обходов по данной граф-схеме (соответственно k образующим), получаем для любой вершины $0 \leqslant x < n$

$$\sum_{i=0}^{k-1} (|c_i| + |c''_i|) \leqslant k\lceil t/2 \rceil + 1 = kT + 1 = 2d + 1.$$

Следовательно, или $D^+(x) \leqslant d$, или $D^-(x) \leqslant d$. Покажем, что в данном графе существует хотя бы одна такая вершина x, что $D(x) = d$. Рассмотрим два возможных варианта.

С л у ч а й (3). Пусть $k \geqslant 3$ — нечётное число и $t \equiv 3 \bmod 4$. В качестве искомой вершины возьмём $x_0 = \dfrac{t+1}{4} \sum\limits_{i=0}^{k-1} t^i$. Имеем $x_0 - n = -\dfrac{t+1}{4} \sum\limits_{i=0}^{k-2} t^i - \dfrac{t+5}{4} t^{k-1}$. Таким образом, $\sum\limits_{i=0}^{k-1} |c_i| = \dfrac{t+1}{4} k = d$ и $\sum\limits_{i=0}^{k-1} |c''_i| = \dfrac{t+1}{4}(k-1) + \dfrac{t+5}{4} = \dfrac{t+1}{4} k + 1 = d + 1$. Рассмотрение длин альтернативных путей из 0 в x_0 показывает, что они не меньше d. Следовательно, $D(x_0) = d$.

Алгоритм 1. Алгоритм преобразования коэффициентов

Вход: параметр t; коэффициенты c_i и c'_i для $i = 0, \ldots, k-1$.
Выход: суммы $|c_i| + |c''_i|$, $i = 0, \ldots, k-1$.

1: $i = -1$.
2: Увеличить i на 1. Положить $c''_i = c'_i$.
3: **Если** $i < k-1$ и $c_i \leqslant 0$, **то**
заменяем c'_i на $c''_i = c''_i + t$. Очевидно, при такой замене сумма $|c''_i| + |c''_{i+1}|$ не увеличивается по сравнению с суммой $|c'_i| + |c'_{i+1}|$ (соответственно в последующем будет замена c'_{i+1} на $c''_{i+1} = c'_{i+1} - 1$). При этом $0 \leqslant c''_i \leqslant \lfloor t/2 \rfloor$ и $|c_i| + |c''_i| = \lceil t/2 \rceil - 1$, перейти в п. 8.
4: **Если** $i = k-1$, **то**
5: из условия $0 \leqslant c_{k-1} \leqslant \lceil t/2 \rceil + 1$ следует $-\lceil t/2 \rceil - 1 \leqslant c''_{k-1} = c'_{k-1} \leqslant 0$ и $|c_{k-1}| + |c''_{k-1}| = \lceil t/2 \rceil + 1$, перейти в п. 15.
6: **Если** $i < k-1$ и $c_i > 0$, **то**
7: $|c'_i| = |c_i - \lceil t/2 \rceil| \leqslant \lfloor t/2 \rfloor$. Тогда $|c_i| + |c''_i| = \lceil t/2 \rceil$, перейти в п. 2.
8: Увеличим i на 1 и положим $c''_i = c'_i - 1$.
9: **Если** $i = k-1$, **то**
10: $|c_{k-1}| + |c''_{k-1}| = \lceil t/2 \rceil + 2$, перейти в п.15.
11: **Если** $c_i > 1$, **то**
12: $|c_i| + |c''_i| = \lceil t/2 \rceil + 1$, перейти в п. 2.
13: **Если** $c_i \leqslant 1$, **то**
14: заменяем c''_i на новое значение $c''_i = c''_i + t$ (соответственно в последующем будет замена c'_{i+1} на $c''_{i+1} = c'_{i+1} - 1$). При этом $-1 \leqslant c''_i \leqslant \lfloor t/2 \rfloor$ и

$$|c_i| + |c''_i| = \begin{cases} \lceil t/2 \rceil - 2, & \text{если } -\lceil t/2 \rceil + 2 \leqslant c_i \leqslant 0, \\ \lceil t/2 \rceil, & \text{если } c_i \in \{-\lceil t/2 \rceil + 1, 1\}, \end{cases}$$

перейти в п.8.
15: Конец алгоритма.

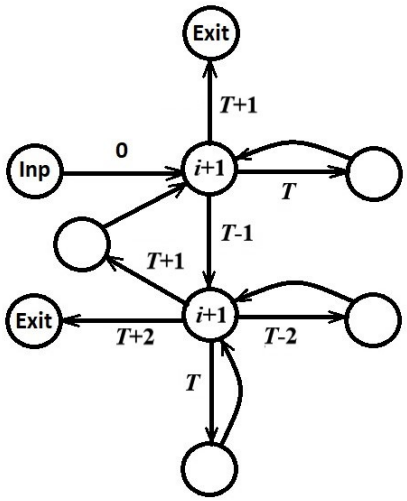

Рис. 1

С л у ч а й (4). Пусть $k \geqslant 4$ — чётное число и $t \equiv 1 \bmod 4$. В качестве искомой вершины возьмём

$$x_0 = \frac{t-1}{2} + t + \frac{t-1}{2}t^2 + t^3 + \ldots + \frac{t-1}{2}t^{k-2} + t^{k-1}.$$

Имеем

$$x_0 - n = -1 - \frac{t-1}{2}t - t^2 - \frac{t-1}{2}t^3 - \ldots - t^{k-2} - \frac{t+1}{2}t^{k-1}.$$

Таким образом, $\sum\limits_{i=0}^{k-1} |c_i| = k/2((t-1)/2+1) = d$ и $\sum\limits_{i=0}^{k-1} |c_i''| = k/2((t-1)/2+1)+1 = d+1$. Длины всех других возможных путей из 0 в x_0 не меньше d. Следовательно, $D(x_0) = d$.

Случай, когда $k \geqslant 4$ — чётное число и $t \equiv 3 \bmod 4$, доказывается аналогично случаю (3). ■

В табл. 2 показан результат сравнения известных семейств мультипликативных циркулянтов с полученными в работе. Здесь n_1, n_2 и n — порядки графов, найденных соответственно посредством теорем 2 [12], 3 [16] и 4.

Т а б л и ц а 2

Сравнение семейств мультипликативных циркулянтов

$k=5$					$k=6$				
d	n_1	n_2	n	t	d	n_1	n_2	n	t
10	682	11 204	13 605	7	9	2 730	11 718	14 843	5
15	18 724	96 630	111 271	11	12	2 730	78 432	95 239	7
20	135 726	433 928	484 553	15	15	149 796	332 150	391 199	9
25	559 240	1 375 610	1 505 931	19	18	149 796	1 062 936	1 223 987	11
30	244 210	3 510 732	3 790 573	23	21	1 628 718	2 815 638	3 186 931	13

Поясним на примерах результат сравнения двух новых семейств с известными семействами мультипликативных циркулянтов.

1. Пусть $k = 5$ и $d = 25$. Тогда в силу теоремы 2 имеем $M(d, 5) \geqslant n = 559\,240$. Теорема 3 даёт следующую оценку: $M(d, 5) \geqslant n = 1\,375\,610$. Новое семейство при том же диаметре даёт оценку $M(d, 5) \geqslant n = 1\,505\,931$. Данный порядок циркулянта достигается при образующих $S = (1, t, t^2, t^3, t^4)$, где $t = 19$.

2. Пусть $k = 6$ и $d = 18$. Тогда в силу теоремы 2 $M(d, 5) \geqslant n = 149\,796$. Теорема 3 даёт следующую оценку: $M(d, 5) \geqslant n = 1\,062\,936$. Новое семейство при диаметре $d = 18$ даёт оценку $M(d, 5) \geqslant n = 1\,223\,987$. Данный порядок циркулянта достигается при образующих $S = (1, t, t^2, t^3, t^4, t^5)$, где $t = 11$.

С помощью специально разработанной компьютерной программы показано, что порядки графов новых семейств мультипликативных циркулянтов являются максимально возможными для исследуемых типов образующих при диаметрах $d \leqslant 22$ и 20 и размерностях соответственно $k = 4$ и 5. Для дальнейшей работы представляет интерес возможность доказательства обобщения данного результата для любых размерностей и диаметров.

Таким образом, полученная в настоящей работе оценка функции $M(d, k)$, подтверждённая конструктивно построением семейств мультипликативных циркулянтных сетей, для размерностей $k \geqslant 5$ (чётных степеней 10 и более) остается пока лучшей известной оценкой.

ЛИТЕРАТУРА

1. *Монахова Э. А.* Структурные и коммуникативные свойства циркулянтных сетей // Прикладная дискретная математика. 2011. № 3. С. 92–115.
2. *Monakhova E. A.* A survey on undirected circulant graphs // Discr. Math., Algorithms and Appl. 2012. No. 4(1). P. 17–47.
3. *Perez-Roses H.* Algebraic and computer-based methods in the undirected degree/diameter problem — a bief survey // Electronic J. Graph Theory and Appl. 2014. No. 2(2). P. 166–190.
4. *Erickson A., Stewart I. A., Navaridas J., and Kiasari A. E.* The stellar transformation: From interconnection networks to datacenter networks // Comput. Networks. 2017. No. 113. P. 29–45.
5. *Wong C. K. and Coppersmith D.* A combinatorial problem related to multimodule memory organizations // J. Assoc. Comput. Mach. 1974. No. 21. P. 392–402.
6. *Monakhova E.* Optimal triple loop networks with given transmission delay: Topological design and routing // Intern. Network Optimization Conf. (INOC'2003), Evry/Paris, France, 2003. P. 410–415.
7. *Dougherty R. and Faber V.* The degree-diameter problem for several varieties of Cayley graphs, 1: The Abelian case // SIAM J. Discrete Math. 2004. No. 17(3). P. 478–519.
8. *Lewis R.* The degree-diameter problem for circulant graphs of degree 8 and 9 // arXiv:1404.3948v1, 2014.
9. *Feria-Puron R., Ryan J., and Perez-Roses H.* Searching for large multi-loop networks // Elec. Notes Disc. Math. 2014. No. 46. P. 233–240.
10. *Feria-Puron R., Perez-Roses H., and Ryan J.* Searching for large circulant graphs // arXiv:1503.07357v1 [math.CO] (25 Mar 2015). P. 31.
11. The Degree/Diameter Problem For Circulant Graphs. `http://combinatoricswiki.org/wiki/The_Degree_Diameter_Problem_for_Circulant_Graphs`.
12. *Chen S. and Jia X. -D.* Undirected loop networks // Networks. 1993. No. 23. P. 257–260.
13. *Parhami B.* Chordal rings based on symmetric odd-radix number systems // Proc. Intern. Conf. on Communications in Computing (Las Vegas, NV, June 27–30). Los Alamitos: IEEE Press, 2005. P. 196–199.
14. *Parhami B.* A class of odd-radix chordal ring networks // The CS'J J. Comput. Sci. Eng. 2006. Vol. 4. No. 2–4. P. 1–9.
15. *Stojmenovic I.* Multiplicative circulant networks. Topological properties and communication algorithms // Discr. Appl. Math. 1997. Vol. 77. P. 281–305.
16. *Monakhova E. A.* On an extremal family of circulant networks // J. Appl. Industr. Math. 2011. No. 5(4). P. 1–7.

REFERENCES

1. *Monakhova E. A.* Strukturnye i kommunikativnye svojstva cirkulyantnyh setej [Structural and communicative properties of circulant networks]. Prikladnaya Diskretnaya Matematika, 2011, no. 3, pp. 92–115. (in Russian)
2. *Monakhova E. A.* A Survey on undirected circulant graphs. Discr. Math., Algorithms and Appl., 2012, no. 4(1), pp. 17–47.
3. *Perez-Roses H.* Algebraic and computer-based methods in the undirected degree/diameter problem — a bief survey. Electronic J. Graph Theory and Appl., 2014, no. 2(2), pp. 166–190.
4. *Erickson A., Stewart I. A., Navaridas J., and Kiasari A. E.* The stellar transformation: From interconnection networks to datacenter networks. Comput. Networks, 2017, no. 113, pp. 29–45.

5. *Wong C. K. and Coppersmith D.* A combinatorial problem related to multimodule memory organizations. J. Assoc. Comput. Mach., 1974, no. 21, pp. 392–402.
6. *Monakhova E.* Optimal triple loop networks with given transmission delay: Topological design and routing. Intern. Network Optimization Conf. (INOC'2003), Evry/Paris, France, 2003, pp. 410–415.
7. *Dougherty R. and Faber V.* The degree-diameter problem for several varieties of Cayley graphs, 1: The Abelian case. SIAM J. Discrete Math., 2004, no. 17(3), pp. 478–519.
8. *Lewis R.* The degree-diameter problem for circulant graphs of degree 8 and 9. arXiv:1404.3948v1, 2014.
9. *Feria-Puron R., Ryan J., and Perez-Roses H.* Searching for large multi-loop networks. Elec. Notes Disc. Math., 2014, no. 46, pp. 233–240.
10. *Feria-Puron R., Perez-Roses H., and Ryan J.* Searching for large circulant graphs. arXiv:1503.07357v1 [math.CO] (25 Mar 2015), p. 31.
11. The Degree/Diameter Problem For Circulant Graphs. `http://combinatoricswiki.org/wiki/The_Degree_Diameter_Problem_for_Circulant_Graphs`.
12. *Chen S. and Jia X. -D.* Undirected loop networks. Networks, 1993, no. 23, pp. 257–260.
13. *Parhami B.* Chordal rings based on symmetric odd-radix number systems. Proc. Intern. Conf. on Communications in Computing (Las Vegas, NV, June 27–30). Los Alamitos, IEEE Press, 2005, pp. 196–199.
14. *Parhami B.* A class of odd-radix chordal ring networks. The CS'J J. Comput. Sci. Eng., 2006, vol. 4, no. 2–4, pp. 1–9.
15. *Stojmenovic I.* Multiplicative circulant networks. Topological properties and communication algorithms. Discr. Appl. Math., 1997, vol. 77, pp. 281–305.
16. *Monakhova E. A.* On an extremal family of circulant networks. J. Appl. Industr. Math., 2011, no. 5(4), pp. 1–7.

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ И ПРОГРАММИРОВАНИЯ

УДК 519.766.2

МИНИМИЗАЦИЯ СИНТАКСИЧЕСКИХ ДИАГРАММ С МНОГОВХОДОВЫМИ КОМПОНЕНТАМИ[1]

Ю. Д. Рязанов

Белгородский государственный технологический университет им. В. Г. Шухова, г. Белгород, Россия

Рассмотрена задача минимизации синтаксических диаграмм. Для её решения диаграммы Вирта (ДВ) преобразуются в синтаксические диаграммы с многовходовыми компонентами (СД), которые по структуре совпадают с ДВ, но отличаются тем, что нетерминалы в нетерминальных вершинах заменяются начальными узлами соответствующих компонент. На множестве узлов СД вводится отношение, обладающее свойством эквивалентности, которое разбивает множество узлов на классы эквивалентности. Доказано, что «стягивание» класса эквивалентности в один узел является эквивалентным преобразованием. Если классу эквивалентности принадлежат узлы различных компонент, то в результате «стягивания» происходит соединение компонент в одну, которая имеет несколько входов. Предложены алгоритмы разбиения множества узлов на классы эквивалентности и построения СД. Приводится пример, показывающий, что построенная по предложенным алгоритмам СД значительно меньше эквивалентной ей ДВ.

Ключевые слова: *формальный язык, синтаксическая диаграмма, отношение эквивалентности, минимизация.*

DOI 10.17223/20710410/41/9

MINIMIZATION OF SYNTAX DIAGRAMS WITH MULTIPORT COMPONENTS

Yu. D. Ryazanov

Belgorod State Technological University named after V. G. Shukhov, Belgorod, Russia

E-mail: razanov.yd@bstu.ru

The minimization problem for syntax diagrams is considered. For this purpose, we transform the Wirth diagrams (WD) into syntax diagrams with multiport components (SD), which are similar to WD in their structure, but differ in the fact that the non-terminals in nonterminal nodes are replaced with the starting nodes of the corresponding components. On the set of SD nodes, we introduce a relation which possesses the equivalence property, dividing the set of nodes into equivalence classes. We prove that uniting an equivalence class into one node is an equivalence transformation. If an equivalence class includes the nodes of various components, then,

[1]Работа поддержана грантом РФФИ № 16-07-00487.

as a result of uniting the class into one node, the components are united into one component, which has several inputs. The algorithms for dividing the set of nodes into equivalence classes and plotting a SD are suggested. The algorithm for dividing the set of nodes into equivalence classes is based on a serial partitioning of the set of nodes into subsets so that non-equivalent nodes fall into different subsets. After partitioning the set of nodes into equivalence classes, an SD is constructed. In the SD construction algorithm, for each equivalence class, the following actions are executed: only one node from the class is left in the SD, the remaining nodes of the class are deleted, and if some arc in the source diagram is going to the deleted node, then it is redirected to one of remaining nodes. We give an example which demonstrates that a SD plotted by the suggested algorithms is considerably smaller than the equivalent WD. The resulting SD, after minimization process, can be used to construct memory efficient programs for formal languages processing.

Keywords: *formal language, syntax diagram, equivalence relation, minimization.*

Введение

Синтаксические диаграммы Вирта представляют собой наглядный, интуитивно понятный графический способ задания синтаксиса языка, они используются для документирования языков программирования [1, 2] и в проектировании трансляторов [3 – 7]. Для построения трансляторов линейной сложности применяются детерминированные синтаксические диаграммы [8, 9], при этом размер транслятора зависит от размера синтаксической диаграммы. Поэтому с целью уменьшения размера транслятора целесообразно использовать «компактные» синтаксические диаграммы.

Характерной особенностью классических синтаксических диаграмм Вирта является то, что каждая компонента имеет ровно один вход и один выход. Эти ограничения позволяют легко строить наглядные и понятные диаграммы для задания контекстно-свободных языков.

В работе рассматриваются синтаксические диаграммы, в которых снято ограничение на количество входов в компонентах. Это позволяет сократить количество компонент и узлов диаграммы. Для сокращения количества компонент и узлов вводится отношение на множестве узлов, обладающее свойством эквивалентности, и предлагается алгоритм, позволяющий разбить множество узлов на классы эквивалентности и на основе этого преобразовать исходную диаграмму в более компактную.

Платой за компактность представления языка в виде синтаксической диаграммы с многовходовыми компонентами является некоторое снижение наглядности по сравнению с классическими диаграммами. Снижение наглядности представления языка в данном случае не является существенным, так как все преобразования могут быть выполнены автоматически по описанным алгоритмам.

В первой части работы поясняются понятия, связанные с диаграммами Вирта (ДВ), необходимые для дальнейшего изложения. Во второй части даётся определение синтаксической диаграммы с многовходовыми компонентами (СД) и способ преобразования ДВ в эквивалентную ей «вырожденную» СД. Далее определяются отношение на множестве узлов СД, названное отношением сильной эквивалентности; преобразование СД, сохраняющее сильную эквивалентность; алгоритм разбиения множества узлов на классы эквивалентности и сокращения СД. Все алгоритмы сопровождаются примерами их применения.

1. Синтаксические диаграммы Вирта

На рис. 1 представлен пример синтаксической диаграммы Вирта. Эта ДВ состоит

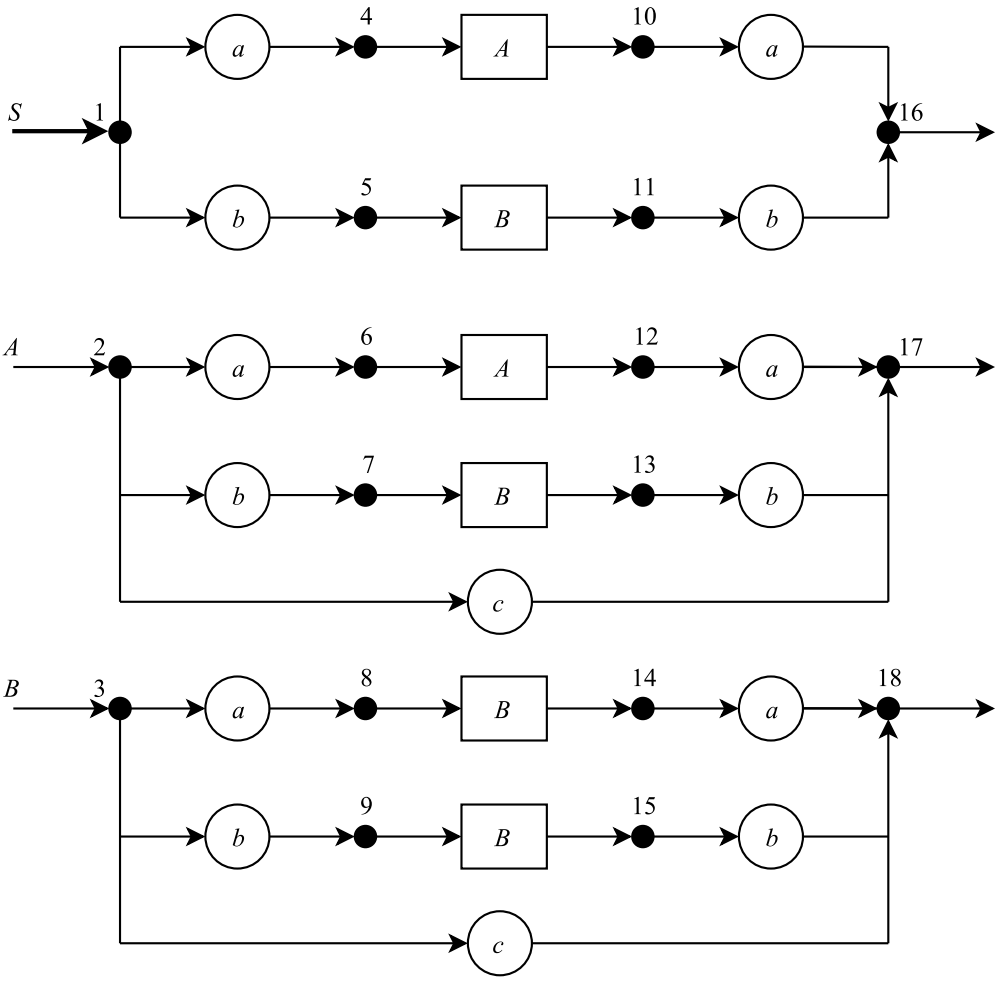

Рис. 1. Синтаксическая диаграмма Вирта

из трёх компонент, соответствующих нетерминалам S, A и B (S — начальный нетерминал). Прямоугольниками изображены нетерминальные вершины, в которые вписаны нетерминалы, кружками — терминальные вершины, в которые вписаны терминалы, жирными точками — узлы, обозначенные натуральными числами. Множество дуг разбивается на четыре подмножества:

1) входящие дуги, которые входят в узлы из одной точки входа;
2) выходящие дуги, которые выходят из узлов и входят в одну точку выхода;
3) внутренние дуги, которые связывают либо узлы с терминальными или нетерминальными вершинами, либо терминальные или нетерминальные вершины с узлами;
4) ε-дуги, связывающие между собой узлы (в примере на рис. 1 таких дуг нет).

Точки входа и выхода на диаграмме не показываются. Узлы, в которые входят входящие дуги, называются начальными (в примере на рис. 1 начальными являются узлы 1, 2 и 3), а узлы, из которых выходят выходящие дуги, — заключительными (в примере на рис. 1 заключительными являются узлы 16, 17 и 18).

С помощью ДВ можно получить (вывести) любую цепочку языка, заданного этой ДВ. Рассмотрим подробно способы получения цепочки языка.

Пусть ДВ G состоит из компонент $G_1, G_2, \ldots, G_{|N|}$, где $|N|$ — мощность множества нетерминалов. Рассмотрим некоторой путь в компоненте G_i от узла u до какого-либо заключительного узла в этой компоненте. Пройдём по нему и в процессе прохождения будем добавлять в изначально пустую цепочку α символы, записанные в терминальных и нетерминальных вершинах. Цепочку α будем называть цепочкой, связывающей узел u с заключительным узлом компоненты G_i. Множество всех цепочек, связывающих узел u с каким-либо заключительным узлом, образует язык $L(u)$. Очевидно, что язык $L(G_i)$ равен объединению всех $L(u_j)$, где u_j — начальный узел компоненты G_i. Вывод цепочки языка, заданного ДВ, можно выполнить по следующему правилу:

1. Записать цепочку, принадлежащую языку $L(G_1)$, где компонента G_1 соответствует начальному нетерминалу.
2. Если цепочка содержит некоторый нетерминал N_i, то заменить его цепочкой языка $L(G_i)$ и выполнить п. 2, иначе — вывод закончить.

Ниже представлен вывод цепочки в ДВ, изображенной на рис. 1:

$$S \Rightarrow aAa \Rightarrow abBba \Rightarrow abcba.$$

Процесс вывода терминальной цепочки в ДВ можно представить и по-другому — как «движение» по дугам от точки входа начальной компоненты к точке выхода. При этом если дуга идёт в терминальную вершину, то вписанный в неё символ добавляем в терминальную цепочку; если в нетерминальную, то переходим в соответствующую компоненту и движемся по ней аналогичным образом до точки выхода, после чего возвращаемся в предыдущую компоненту и продолжаем движение. После прохождения выходной дуги начальной компоненты в терминальную цепочку добавляем концевой маркер ($\dashv$) и вывод заканчивается.

Символ x, который может быть добавлен в терминальную цепочку непосредственно после прохождения дуги e, принадлежит множеству выбора дуги e ($x \in$ ВЫБОР(e)).

В множестве ДВ можно выделить класс детерминированных ДВ. В детерминированной ДВ каждая компонента имеет только один начальный узел, не содержит ε-дуг и каждый её узел детерминированный. Узел является детерминированным, если множества выбора любых двух дуг, выходящих из него, не пересекаются. Детерминированные ДВ являются основой для построения программ-распознавателей линейной сложности [8, 9], поэтому в дальнейшем будем рассматривать только детерминированные ДВ.

2. Синтаксические диаграммы с многовходовыми компонентами

Синтаксическую диаграмму с многовходовыми компонентами (СД) определим восьмеркой $R = (T, U, U', U'', u_0, G, F_T, F_U)$, где

— T — конечное множество терминалов;
— U — конечное множество узлов;
— U' — конечное множество начальных узлов, $U' \subseteq U$;
— U'' — конечное множество заключительных узлов, $U'' \subseteq U$;
— u_0 — стартовый узел, $u_0 \in U'$;
— $G = (V, E)$ — ориентированный граф, $V = V_T \cup V_N \cup U$, V_T — множество терминальных вершин, V_N — множество нетерминальных вершин; $E = E_1 \cup E_2$, $E_1 \subseteq U \times (V_T \cup V_N)$ — множество дуг, выходящих из узлов и входящих в терминальные или нетерминальные вершины; $E_2 \subseteq (V_T \cup V_N) \times U$ — множество дуг, выходящих из терминальных или нетерминальных вершин и входящих в узлы;

— $F_T : V_T \to T$ — отображение множества терминальных вершин в множество терминалов;
— $F_U : V_N \to U'$ — отображение множества нетерминальных вершин в множество начальных узлов.

Терминальная вершина изображается на диаграмме кружком, в который вписан терминал в соответствии с отображением F_T. Нетерминальная вершина изображается прямоугольником, в который вписан узел из множества начальных узлов в соответствии с отображением F_U. Узел изображается жирной точкой, которая отмечается соответствующим номером. Начальные узлы отмечаются входящей стрелочкой, стартовый узел — жирной входящей стрелочкой, заключительные — выходящей стрелочкой. Дуга может выходить из терминальной или нетерминальной вершины и входить в узел, или выходить из узла и входить в терминальную или нетерминальную вершину. В каждую терминальную и нетерминальную вершину входит только одна дуга; из них выходит только одна дуга. На количество дуг, входящих в узлы и выходящих из узлов, ограничений нет.

Определим способ получения цепочки языка, заданного СД.

Будем говорить, что цепочка α, состоящая из терминалов и/или начальных узлов, связывает узел u СД с заключительным узлом u_k, если её можно получить, «двигаясь» в СД от узла u к узлу u_k и выписывая из вершин по пути символы (терминалы или начальные узлы) в изначально пустую цепочку. Множество всех цепочек, связывающих узел u с заключительными узлами, обозначим $L(u)$. Чтобы получить цепочку языка, заданного СД со стартовым узлом u_0, возьмём цепочку, принадлежащую $L(u_0)$. Если она содержит некоторый начальный узел u_i, заменим его на цепочку из множества $L(u_i)$. Если новая цепочка содержит начальный узел, то аналогичные действия повторяем. Полученная таким образом цепочка, не содержащая начальных узлов, принадлежит языку, заданному СД.

ДВ можно преобразовать в эквивалентную ей СД, заменив записанный в каждой нетерминальной вершине нетерминал на его начальный узел и определив начальный узел начальной компоненты (соответствующей начальному нетерминалу) стартовым. На рис. 2 представлена СД, полученная из ДВ (рис. 1).

Сравнивая способы получения цепочек языка по ДВ и СД, можно сделать вывод, что язык, заданный ДВ, равен языку, заданному СД, полученной из ДВ описанным способом.

Процесс получения одной из цепочек языка, заданного СД на рис. 2, можно представить следующим образом:

$$1 \Rightarrow a2a \Rightarrow ab3ba \Rightarrow abcba.$$

Такой способ преобразования позволяет получить «вырожденную» СД, в которой каждая компонента имеет один вход. Преобразования, описанные далее, позволят получить СД с многовходовыми компонентами и меньшим количеством компонент.

Для СД точно так же, как для ДВ, определяются понятия множество выбора дуги и детерминированность. В дальнейшем будем рассматривать СД, полученные из детерминированных ДВ, следовательно, СД тоже будут детерминированными.

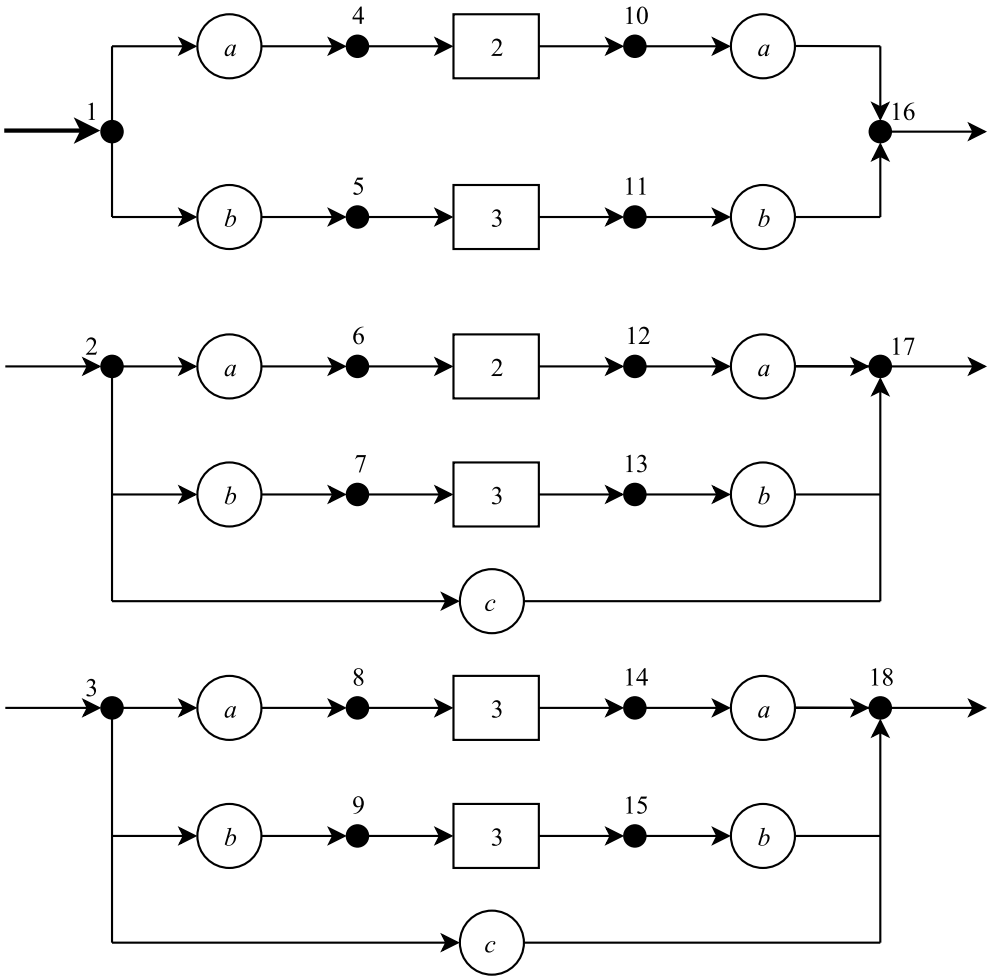

Рис. 2. Синтаксическая диаграмма с многовходовыми компонентами

3. Отношение сильной эквивалентности

Для определения отношения сильной эквивалентности на множестве узлов введём следующие обозначения:

1) (u_i, t, u_j) — указывает на то, что в СД существует путь длины два из узла u_i в узел u_j, дуга из узла u_i идёт в вершину с терминалом t, а из неё идёт дуга в узел u_j;
2) (u_i, M, u_k, u_j) — указывает на то, что в СД существует путь длины два из узла u_i в узел u_j, дуга из узла u_i с множеством выбора M идёт в вершину с узлом u_k, а из неё идёт дуга в узел u_j.

Пара узлов (u_i, u'_i) принадлежит отношению E°, если выполнены следующие условия:

1) путь (u_i, t, u_j) существует тогда и только тогда, когда существует путь (u'_i, t, u'_j) и $(u_i, u'_j) \in E^\circ$, где t принадлежит множеству терминалов;
2) путь (u_i, M, u_k, u_j) существует тогда и только тогда, когда существует путь (u'_i, M, u'_k, u'_j), $(u_j, u'_j) \in E^\circ$ и $(u_k, u'_k) \in E^\circ$, где u_k и u'_k принадлежат множеству начальных узлов;
3) $u_i \in U''$ тогда и только тогда, когда $u'_i \in U''$.

Из определения отношения E° следует, что цепочка α связывает узел u_i с заключительным узлом тогда и только тогда, когда цепочка α' связывает узел u'_i с заключительным узлом, где α и α' — цепочки одинаковой длины и i-й символ цепочки α либо равен i-му символу цепочки α' ($\alpha_i = \alpha'_i$), либо пара начальных узлов (α_i, α'_i) принадлежит отношению E°. Цепочки α и α' назовём сильно эквивалентными, а отношение E° — отношением сильной эквивалентности.

На множествах цепочек введём следующие операции сравнения:

1) $L_1 \subseteq^\circ L_2$ — истина, если для каждой цепочки множества L_1 существует сильно эквивалентная цепочка из множества L_2, иначе — ложь;
2) $L_1 =^\circ L_2$ тогда и только тогда, когда $L_1 \subseteq^\circ L_2$ и $L_2 \subseteq^\circ L_1$.

Операции $=^\circ$ и $\subseteq^\circ$ будем называть соответственно равенством и включением. Отношение E° можно теперь определить так:

$$E^\circ = \{(u_i, u'_i) : L(u_i) =^\circ L(u'_i)\}.$$

Очевидно, что

1) если $L_1 = L_2$, то $L_1 =^\circ L_2$;
2) если $L_1 \subseteq L_2$, то $L_1 \subseteq^\circ L_2$;
3) $L_1 \cup L_2 \subseteq^\circ L_3$ тогда и только тогда, когда $L_1 \subseteq^\circ L_3$ и $L_2 \subseteq^\circ L_3$.

Для множеств L_1, L_2, L_3 и L_4, таких, что $L_1 =^\circ L_3$ и $L_2 =^\circ L_4$, верно:

1) $L_1 \cup L_2 =^\circ L_3 \cup L_4$;
2) $L_1 L_2 =^\circ L_3 L_4$;
3) $L_1^* =^\circ L_3^*$.

Рассмотрим две СД $R_1 = (T, U^1, U'^1, U''^1, u_0^1, G^1, F_T^1, F_U^1)$ и $R_2 = (T, U^2, U'^2, U''^2, u_0^2, G^2, F_T^2, F_U^2)$, такие, что $(u_0^1, u_0^2) \in E^\circ$. Докажем, что $L(R_1) = L(R_2)$, т. е. эти СД определяют один и тот же язык.

Пусть на первом шаге вывода в СД R_1 получена промежуточная цепочка α, принадлежащая языку $L(u_0^1)$. Из $L(u_0^1) =^\circ L(u_0^2)$ следует, что на первом шаге вывода в СД R_2 можно получить промежуточную цепочку α', сильно эквивалентную цепочке α. Далее на каждом шаге вывода в СД R_1 и R_2 будем получать сильно эквивалентные цепочки вплоть до терминальной цепочки. Терминальные сильно эквивалентные цепочки равны.

Пусть на первом шаге вывода в СД R_2 получена промежуточная цепочка α, принадлежащая языку $L(u_0^2)$. Из $L(u_0^1) =^\circ L(u_0^2)$ следует, что на первом шаге вывода в СД R_1 можно получить промежуточную цепочку α', сильно эквивалентную цепочке α. Далее на каждом шаге вывода в СД R_1 и R_2 будем получать сильно эквивалентные цепочки вплоть до терминальной цепочки. Терминальные сильно эквивалентные цепочки равны.

В любом случае терминальные цепочки, которые можно получить при выводе в одной СД, можно получить и в другой СД.

4. Преобразование, сохраняющее сильную эквивалентность

Допустим, что над СД $R_1 = (T, U^1, U'^1, U''^1, u_0^1, G^1, F_T^1, F_U^1)$ выполнено преобразование и в результате получена СД $R_2 = (T, U^2, U'^2, U''^2, u_0^2, G^2, F_T^2, F_U^2)$, такая, что $(u_0^1, u_0^2) \in E^\circ$. В этом случае будем говорить, что над СД R_1 выполнено преобразование, сохраняющее сильную эквивалентность. Результат преобразования — СД R_2, которая эквивалентна СД R_1. Рассмотрим преобразование СД, сохраняющее сильную эквивалентность.

Пусть в СД $R_1 = (T, U, U'^1, U'', u_0^1, G^1, F_T, F_U)$: $U'^1 = \{u_0, u_1, \dots u_f, \dots, u_n\}$ — множество начальных узлов; $(u_i, u_i') \in E^\circ$; узлы u_i и u_i' не являются начальными. Преобразование заключается в том, что все дуги, входящие в узел u_i, перенаправляются в узел u_i'. В результате преобразования получается СД $R_2 = (T, U, U'^2, U'', u_0^2, G^2, F_T, F_U)$.

Докажем, что это преобразование сохраняет сильную эквивалентность. Из множества начальных узлов U'^1 выбираем произвольно узел u_f. Введём следующие обозначения:

1) $L_1(u_f, u_i)$ — множество цепочек, связывающих начальный узел u_f с узлом u_i в исходной СД;
2) $L_1(u_i, U'')$ — множество цепочек, связывающих узел u_i с заключительными узлами U'' в СД R_1;
3) $L_1(u_f, u_i, U'')$ — множество цепочек, связывающих начальный узел u_f с заключительными узлами U'', соответствующих путям, проходящим через узел u_i в СД R_1; $L_1(u_f, u_i, U'') = L_1(u_f, u_i)L_1(u_i, U'')$ — множество $L_1(u_f, u_i, U'')$ равно конкатенации множеств $L_1(u_f, u_i)$ и $L_1(u_i, U'')$;
4) $L_1(u_f, \overline{u_i}, U'')$ — множество цепочек, связывающих начальный узел u_f с заключительными узлами U'', соответствующих путям, не проходящим через узел u_i в СД R_1;
5) $L_1(u_f, U'')$ — множество цепочек, связывающих начальный узел u_f с заключительными узлами U'' в СД R_1; $L_1(u_f, U'') = L_1(u_f, u_i, U'') \cup L_1(u_f, \overline{u_i}, U'') = L_1(u_f, u_i)L_1(u_i, U'') \cup L_1(u_f, \overline{u_i}, U'')$;
6) $L_2(u_f, u_i')$ — множество цепочек, связывающих начальный узел u_f с узлом u_i' в СД R_2; $L_2(u_f, u_i') = L_1(u_f, u_i') \cup L_1(u_f, u_i)$;
7) $L_2(u_f, u_i', U'')$ — множество цепочек, связывающих начальный узел u_f с заключительными узлами U'', соответствующих путям, проходящим через узел u_i' в СД R_2;

$$L_2(u_f, u_i', U'') = L_2(u_f, u_i')L_1(u_i', U'') = (L_1(u_f, u_i') \cup L_1(u_f, u_i))L_1(u_i', U'') = \\ = L_1(u_f, u_i')L_1(u_i', U'') \cup L_1(u_f, u_i)L_1(u_i', U'');$$

8) $L_2(u_f, U'')$ — множество цепочек, связывающих начальный узел u_f с заключительными U'' в СД R_2;

$$L_2(u_f, U'') = L_2(u_f, u_i', U'') \cup L_1(u_f, \overline{u_i}, U'') = \\ = L_1(u_f, u_i')L_1(u_i', U'') \cup L_1(u_f, u_i)L_1(u_i', U'') \cup L_1(u_f, \overline{u_i}, U'').$$

Для доказательства того, что рассматриваемое преобразование сохраняет сильную эквивалентность, необходимо доказать, что $L_1(u_f, U'') =^\circ L_2(u_f, U'')$, т. е.

$$L_1(u_f, u_i)L_1(u_i, U'') \cup L_1(u_f, \overline{u_i}, U'') =^\circ \\ =^\circ L_1(u_f, u_i')L_1(u_i', U'') \cup L_1(u_f, u_i)L_1(u_i', U'') \cup L_1(u_f, \overline{u_i}, U'').$$

В свою очередь, $L_1(u_f, U'') =^\circ L_2(u_f, U'')$, если имеют место включения 1) $L_1(u_f, U'') \subseteq {}^\circ L_2(u_f, U'')$ и 2) $L_2(u_f, U'') \subseteq {}^\circ L_1(u_f, U'')$.

1. $L_1(u_f, U'') \subseteq^\circ L_2(u_f, U'')$, если

$$L_1(u_f, u_i)L_1(u_i, U'') \subseteq^\circ L_2(u_f, U'') \quad \text{и} \quad L_1(u_f, \overline{u_i}, U'') \subseteq {}^\circ L_2(uf, U'').$$

Истинность первого условия следует из того, что $L_1(u_f, u_i)L_1(u_i', U'') \subseteq^\circ L_2(u_f, U'')$ по определению и $L_1(u_i, U'') =^\circ L_1(u_i', U'')$, так как $(u_i, u_i') \in E^\circ$.

Истинность второго условия следует из того, что любая цепочка из множества $L_1(u_f, \overline{u_i}, U'')$ включена либо в $L_1(u_f, u'_i)L_1(u'_i, U'')$, либо в $L_1(u_f, \overline{u_i}, U'')$, т. е. она соответствует либо пути, проходящему через узел u'_i, либо пути, не проходящему через узел u'_i.

2. $L_2(u_f, U'') \subseteq^\circ L_1(u_f, U'')$, если

$$L_1(u_f, u'_i)L_1(u'_i, U'') \subseteq^\circ L_1(u_f, U''), \quad L_1(u_f, u_i)L_1(u'_i, U'') \subseteq^\circ L_1(u_f, U'')$$
$$\text{и} \quad L_1(u_f, \overline{u_i}, U'') \subseteq^\circ L_1(u_f, U'').$$

Истинность первого условия следует из того, что любая цепочка из множества $L_1(u_f, u'_i)L_1(u'_i, U'')$ включена либо в $L_1(u_f, u_i)L_1(u_i, U'')$, либо в $L_1(u_f, \overline{u_i}, U'')$, т. е. она соответствует либо пути, проходящему через узел u_i, либо пути, не проходящему через узел u_i.

Истинность второго условия следует из того, что $L_1(u_f, u_i)L_1(u_i, U'') \subseteq L_1(u_f, U'')$ по определению и $L_1(u_i, U'') =^\circ L_1(u'_i, U'')$, так как $(u_i, u'_i) \in E^\circ$.

Истинность третьего условия следует из того, что любая цепочка из множества $L_1(u_f, \overline{u_i}, U'')$ включена либо в $L_1(u_f, u_i)L_1(u_i, U'')$, либо в $L_1(u_f, \overline{u_i}, U'')$, т. е. она соответствует либо пути, проходящему через узел u_i, либо пути, не проходящему через узел u_i.

Таким образом, $L_1(u_f, U'') =^\circ L_2(u_f, U'')$, следовательно, рассматриваемое преобразование сохраняет сильную эквивалентность.

После выполнения преобразования в узел u_i не будет входить ни одна дуга. Он станет недостижимым из начальных узлов, поэтому его нужно удалить вместе с выходящими дугами, вершинами (терминальными и нетерминальными), в которые входят эти дуги, и дугами, выходящими из удаляемых терминальных и нетерминальных вершин. В результате выполнения описанных действий может получиться СД с узлами, в которые не входят дуги. Исключение недостижимых узлов, дуг, терминальных и нетерминальных вершин необходимо повторять, пока это возможно.

Если узел u_i начальный или стартовый, то узел u'_i нужно сделать соответственно начальным или стартовым. Узел u_i будет недостижимым, и его нужно удалить по описанным выше правилам. После этого некоторые нетерминальные вершины будут содержать несуществующий узел u_i. В таких вершинах его нужно заменить на u'_i.

Рассмотренное преобразование можно назвать «стягиванием» сильно эквивалентных узлов в один узел. Очевидно, что таким образом можно «стянуть» в один узел класс сильно эквивалентных узлов, что уменьшит размер СД. Если при этом «стягиваемые» узлы принадлежат различным компонентам, то происходит соединение компонент в одну, в которой начальными узлами будут начальные узлы соединяемых компонент. Таким образом появляются компоненты со многими входами.

5. Разбиение множества узлов на классы эквивалентности по отношению сильной эквивалентности

Отношение сильной эквивалентности определяет разбиение множества узлов СД на классы эквивалентности. Применим метод последовательного разбиения.

Сначала разобьём узлы СД на подмножества так, чтобы заключительные узлы попали в одно подмножество, а все остальные — в другое. Для СД на рис. 2 это будут подмножества $Q_1 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$ и $Q_2 = \{16, 17, 18\}$, которые образуют разбиение $R = \{Q_1, Q_2\}$.

Узлы, принадлежащие различным подмножествам, явно неэквивалентны (для них нарушено третье условие сильной эквивалентности); два узла из одного подмножества

могут быть неэквивалентными, так как при разбиении мы не учитывали первое и второе условия принадлежности пары узлов отношению сильной эквивалентности.

Для учёта этих условий построим таблицу T, в которой столбцы соответствуют узлам СД, а строки — терминалам. Клетку таблицы в строке t и столбце u будем обозначать $T[t, u]$.

Пути (u_i, t, u_j) соответствует клетка $T[t, u_i]$. Если $u_j \in Q_r$, то в клетку $T[t, u_i]$ запишем r. Пути (u_i, M, u_k, u_j) соответствуют клетки $T[t, u_i]$, такие, что $t \in M$. Если $u_j \in Q_r$ и $u_k \in Q_s$, то в клетку $T[t, u_i]$ запишем (r, s). Если узлы u_i и u'_i принадлежат одному подмножеству и $T[t, u_i] \neq T[t, u'_i]$, то это говорит о том, что пара узлов u_i и u'_i не принадлежит отношению сильной эквивалентности и узлы u_i и u'_i нужно включить в разные подмножества нового разбиения R'.

Анализируя таблицу T, формируем новое разбиение R'. Подмножество Q'_k разбиения R' формируется из элементов некоторого подмножества Q_k разбиения R: Q'_k является максимальным по мощности подмножеством множества Q_k, таким, что для каждой пары узлов u_i и u'_i, принадлежащих подмножеству Q'_k, для всех строк t таблицы верно, что $T[t, u_i] = T[t, u'_i]$. Если $R' \neq R$, то полагаем $R = R'$, заново строим таблицу T по описанным правилам и формируем новое разбиение. Если $R' = R$, то R' — множество классов эквивалентности.

Первая таблица T для СД на рис. 2 представлена в табл. 1.

Таблица 1

Терминалы	Q_1															Q_2		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
a	1	1	1	1 1	1 1	1 1	1 1	1 1	1 1	2		2		2				
b	1	1	1	1 1	1 1	1 1	1 1	1 1	1 1		2		2		2			
c		2	2	1 1	1 1	1 1	1 1	1 1	1 1									

По табл. 1 видно, что подмножество Q_1 разбивается на подмножества $\{1\}$, $\{2, 3\}$, $\{4, 5, 6, 7, 8, 9\}$, $\{10, 12, 14\}$, $\{11, 13, 15\}$, а подмножество Q_2 не разбивается. В результате получаем разбиение $\{\{1\}, \{2, 3\}, \{4, 5, 6, 7, 8, 9\}, \{10, 12, 14\}, \{11, 13, 15\}, \{16, 17, 18\}\}$. По нему строим новую таблицу T (табл. 2).

Таблица 2

Терминалы	Q_1	Q_2		Q_3						Q_4			Q_5			Q_6		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
a	3	3	3	4 2	5 2	4 2	5 2	4 2	5 2	6	6	6						
b	3	3	3	4 2	5 2	4 2	5 2	4 2	5 2				6	6	6			
c		6	6	4 2	5 2	4 2	5 2	4 2	5 2									

По табл. 2 видно, что подмножество $Q_3 = \{4, 5, 6, 7, 8, 9\}$ разбивается на подмножества $\{4, 6, 8\}$ и $\{5, 7, 9\}$, а остальные подмножества не разбиваются. В результате получаем разбиение $\{\{1\}, \{2, 3\}, \{4, 6, 8\}, \{5, 7, 9\}, \{10, 12, 14\}, \{11, 13, 15\}, \{16, 17, 18\}\}$. По нему строим табл. 3.

Таблица 3

Терминалы	Q_1	Q_2		Q_3			Q_4			Q_5			Q_6			Q_7		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
a	3	3	3	5 2	5 2	5 2	6 2	6 2	6 2	7	7	7						
b	4	4	4	5 2	5 2	5 2	6 2	6 2	6 2				7	7	7			
c		7	7	5 2	5 2	5 2	6 2	6 2	6 2									

В табл. 3 нет различных столбцов, соответствующих узлам, принадлежащим одному подмножеству. Поэтому полученное разбиение является разбиением на классы эквивалентности.

6. Минимизация синтаксической диаграммы с многовходовыми компонентами на основе отношения сильной эквивалентности

Используя отношение сильной эквивалентности на множестве узлов СД R_1, можно построить СД R_2, эквивалентную R_1, которая будет иметь не больше узлов и компонент, чем R_1, а компоненты R_2 могут иметь большее количество входов, чем компоненты R_1. Сокращение количества узлов достигается за счёт «стягивания» всех узлов, принадлежащих одному классу сильной эквивалентности, в один узел.

Для преобразования СД R_1 в СД R_2 удобно R_1 представить в виде таблицы и преобразовать её в таблицу R_2 (СД R_2 будем называть сокращённой). Столбцы таблицы соответствуют узлам СД, а строки — терминалам и начальным узлам, которые встречаются в нетерминальных вершинах. Пути (u_i, x, u_j) соответствует клетка $T[x, u_i]$, в которой записан узел u_j. Начальные узлы отметим стрелкой, стартовый узел — жирной стрелкой, а заключительные — символом «1». Таблица СД (рис. 2) представлена в табл. 4.

Таблица 4

Таблица синтаксической диаграммы с многовходовыми компонентами

Терминалы и начальные узлы	↓	↓	↓													1	1	1
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
a	4	6	8							16		17		18				
b	5	7	9								16		17		18			
c		17	18															
2				10		12												
3					11		13	14	15									

Отношение сильной эквивалентности определяет разбиение множества узлов на классы $\{\{1\}, \{2, 3\}, \{4, 6, 8\}, \{5, 7, 9\}, \{10, 12, 14\}, \{11, 13, 15\}, \{16, 17, 18\}\}$. Возьмём по одному узлу из каждого класса и получим множество узлов $\{1, 2, 4, 5, 10, 11, 16\}$ сокращённой СД. В таблице СД (табл. 4) оставим только те столбцы, которые соответствуют узлам сокращённой СД. Если строка таблицы отмечена узлом, не принадлежащим множеству узлов сокращённой СД, то отметим её сильно эквивалентным ему узлом сокращённой СД. Строки, отмеченные одинаковыми узлами, объединим в одну. Если в клетке полученной таблицы встретится узел, который не принадлежит множеству узлов сокращённой СД, то его нужно заменить на сильно эквивалентный ему

узел сокращённой СД. Если узел сокращённой СД сильно эквивалентен начальному (стартовому) узлу исходной СД, то нужно сделать его начальным (стартовым).

Таблица сокращённой СД, эквивалентной СД на рис. 2, приведена в табл. 5, её графическое представление изображено на рис. 3. Эта СД содержит значительно меньше узлов, чем исходная (рис. 2), и одну компоненту с двумя входами.

Таблица 5

Таблица сокращённой СД

Терминалы и начальные узлы	↓	↓					1
	1	2	4	5	10	11	16
a	4	4			16		
b	5	5				16	
c		16					
2			10	11			

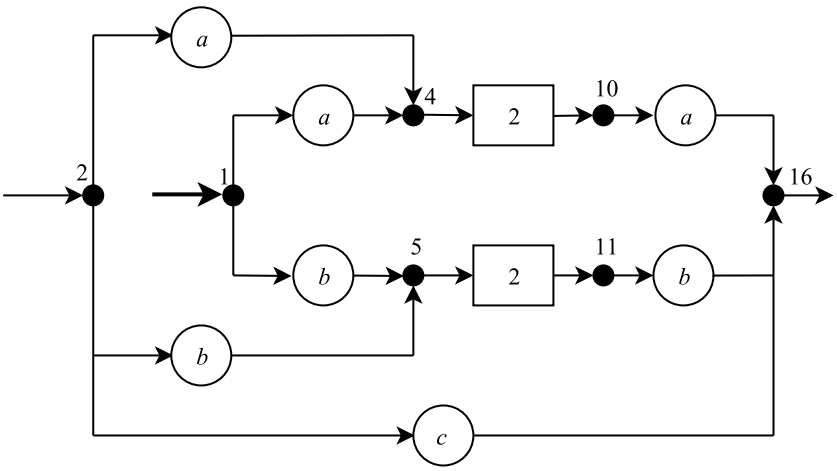

Рис. 3. Сокращённая СД

Заключение

Предложенный метод минимизации позволяет преобразовать диаграмму Вирта в более компактную синтаксическую диаграмму с многовходовыми компонентами. Такая диаграмма может быть использована для построения эффективных по памяти программ-распознавателей линейной сложности.

ЛИТЕРАТУРА

1. *Jensen K. and Wirth N.* Pascal User Manual and Report. N.Y.: Springer Verlag, 1975. 167 p.
2. *Jensen K. and Wirth N.* Pascal User Manual and Report. Berlin, Heidelberg: Springer Verlag, 1974. 170 p.
3. *Легалов А. И., Швец Д. А., Легалов И. А.* Формальные языки и трансляторы. Красноярск: Сибирский федеральный университет, 2007. 213 с.
4. *Карпов Ю. Г.* Теория и технология программирования. Основы построения трансляторов. СПб.: БХВ-Петербург, 2005. 272 с.
5. *Свердлов С. З.* Методы трансляции. Вологда: ВоГУ, 2016. 235 с.
6. *Свердлов С. З.* Конструирование компиляторов. Saarbruken: Lap Lambert, 2015. 571 с.

7. *Мартыненко Б. К.* Синтаксические диаграммы Н. Вирта и граф-схемы в Syntax-технологии // Компьютерные инструменты в образовании. 2014. № 2. С. 3–19.
8. *Рязанов Ю. Д., Севальнева М. Н.* Анализ синтаксических диаграмм и синтез программ-распознавателей линейной сложности // Научные ведомости БелГУ. Сер. История. Политология. Экономика. Информатика. 2013. № 8. С. 128–136.
9. *Поляков В. М., Рязанов Ю. Д.* Алгоритм построения нерекурсивных программ-распознавателей линейной сложности по детерминированным синтаксическим диаграммам // Вестник БГТУ им. В. Г. Шухова. 2013. № 6. С. 194–199.

REFERENCES

1. *Jensen K. and Wirth N.* Pascal User Manual and Report. New York, Springer Verlag, 1975. 167 p.
2. *Jensen K. and Wirth N.* Pascal User Manual and Report. Berlin, Heidelberg, Springer Verlag, 1974. 170 p.
3. *Legalov A. I., Shvets D. A. and Legalov I. A.* Formal'nyye yazyki i translyatory [Formal Languages and Translators]. Siberian Federal University, Krasnoyarsk, 2007. 213 p. (in Russian)
4. *Karpov Yu. G.* Teoriya i tekhnologiya programmirovaniya. Osnovy postroyeniya translyatorov [The Theory and Technology of Programming. Fundamentals of Translators]. Saint-Petersburg, BHV-Petersburg Publ., 2005. 272 p. (in Russian)
5. *Sverdlov S. Z.* Metody translyatsii [Methods of Translation]. Vologda, VSU Publ., 2016. 235 p. (in Russian)
6. *Sverdlov S. Z.* Konstruirovaniye kompilyatorov [Compiler Design]. Saarbruken, Lap Lambert, 2015. 571 p. (in Russian)
7. *Martynenko B. K.* Sintaksicheskie diagrammy N. Virta i graf-shemy v syntax-tehnologii [Syntactic charts and graph-schemes in the SYNTAX-Technology]. Computer Tools in Education, 2014, no. 2, pp. 3–19. (in Russian)
8. *Ryazanov Yu. D. and Seval'neva M. N.* Analiz sintaksicheskih diagramm i sintez programm-raspoznavatelej linejnoj slozhnosti [The analysis of syntax diagrams and automatic generation of linear-time programs recognizer]. Belgorod State University Scientific Bulletin. History. Political science. Economics. Information technologies, 2013, no. 8, pp. 128–136. (in Russian)
9. *Polyakov V. M. and Ryazanov Yu. D.* Algoritm postroeniya nerekursivnyh programm-raspoznavatelej linejnoj slozhnosti po determinirovannym sintaksicheskim diagrammam [Algorithm for not recursive linear-time programs recognizer design from deterministic syntax diagrams]. Bulletin of BSTU named after V. G. Shukhov, 2013, no. 6, pp. 194–199. (in Russian)

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ В ДИСКРЕТНОЙ МАТЕМАТИКЕ

УДК 007.52

АРХИТЕКТУРА НЕЙРОННОЙ СЕТИ С ПОПАРНО ПОСЛЕДОВАТЕЛЬНЫМ РАЗДЕЛЕНИЕМ ОБРАЗОВ

П. Ш. Гейдаров

Институт систем управления НАН Азербайджана, г. Баку, Азербайджанская Республика

На основе архитектуры нейронной сети, реализующей метод ближайшего соседа, рассматривается архитектура нейронной сети с попарно последовательным разделением образов без использования аналитических выражений и набора выбранных эталонов. Изучаются возможности данной архитектуры. Показано, что такая архитектура может быть применена к задаче распознавания с очень большим количеством образов. Предлагаемая нейронная сеть отличается простой и понятной архитектурой, возможностью простого обучения нейронной сети с добавлением в неё новых распознаваемых образов без необходимости изменения предыдущих настроек сети.

Ключевые слова: *архитектуры нейронных сетей, определяемые нейронные сети, нейрокопьютер, сверточные сети, алгоритмы обучения нейронных сетей.*

DOI 10.17223/20710410/41/10

THE ARCHITECTURE OF A NEURAL NETWORK WITH A SEQUENTIAL DIVISION OF IMAGES INTO PAIRS

P. Sh. Geidarov

Institute of Control Systems of ANAS, Baku, Republic of Azerbaijan

E-mail: plbaku2010@gmail.com

The known classical architectures of neural networks have many weak properties and high limitations such as great difficulties in choosing proper parameters (the numbers of neurons, connections, layers), in learning and in expanding a learned network and some others. In this paper, to overcome these shortcomings, we consider the architecture of a neural network in which the recognition is performed in pairs. The neural network is created on the basis of the neural network architecture that implements the method of the nearest neighbor, without using analytical expressions and a set of selected samples. The paper shows that such an architecture can be applied to the recognition problem with a very large number of images (classes). At the same time, the proposed neural network has a simple architecture, with the possibility of simpler learning of the neural network. The neural network architecture allows to add new recognizable images to the neural network without changing the previous network settings.

Keywords: *neural network architectures, neural networks, neurocomputer, convolutional networks, neural network training algorithms.*

Введение

Создание простых и понятных архитектур нейронных сетей с прозрачной архитектурой, способных решать сложные задачи, по-прежнему является актуальной задачей. Известные в настоящее время классические архитектуры нейронных сетей [1 – 3] имеют ряд слабых сторон и ограничений, среди которых такие, как:

1. Сложность выбора структуры архитектуры нейронной сети и её параметров (количества нейронов, связей, слоёв). Эти параметры выбираются либо эмпирически, либо при помощи приближённых подходов и рекомендаций, либо имеют весьма сложную многослойную форму, как, например, для сетей с глубоким обучением [4 – 6].
2. Сложность обучения нейронной сети с большим количеством распознаваемых образов. Чем больше количество распознаваемых образов, тем сложнее обучаются нейронные сети и тем хуже, как правило, результат обучения. Это приводит к ограничению возможного количества образов, используемых в задачах распознавания с применением нейронных сетей.
3. Сложность расширения уже обученной нейронной сети и добавления к ней новых образов, которые часто требуют переобучения сети с полным или частичным изменением предыдущих весовых и пороговых значений.

В данной работе рассматривается возможность создания прозрачной архитектуры нейронной сети на основе нейронной сети, реализующей метрические методы распознавания, что позволит обойти приведённые трудности.

В работах [7, 8] рассмотрены архитектуры нейронных сетей, реализованные на основе метрических методов распознавания (НСММР) [9]. В [10] выполнено преобразование этих сетей в многослойный персептрон с полной системой связи, которое показало, что данные сети являются частным случаем многослойного персептрона. Но при этом в отличие от классических схем многослойного персептрона структура и параметры НСММР, такие, как количество нейронов, связей, слоёв, а также значения весов и порогов вычисляются аналитически, исходя из начальных условий задачи — количества распознаваемых образов, эталонов, используемых метрических выражений. В частности, на рис. 1 приведена схема трёхслойной нейронной сети прямого распространения с пороговой функцией активации, реализующей метод ближайшего соседа [7, 10].

В качестве соседних элементов метода ближайшего соседа используется набор отобранных эталонов из обучающей выборки. Набор эталонов может отбираться как интуитивно, так и при помощи алгоритма обучающего отбора [7].

Структура данной сети (количество нейронов, слоёв и связей) строго определяется согласно схеме на рис. 2. Каждый нейрон первого слоя выполняет попарное сравнение изображений двух эталонов. При этом значения весов нейронов первого слоя определяются аналитически на основе метрических выражений, например для рис. 2, *б* по выражению

$$w_{c,r}^{(1)} = d_1^2 - d_2^2 = \left((c_1 - c_p)^2 + (r_1 - r_p)^2\right) - \left((c_2 - c_p)^2 + (r_2 - r_p)^2\right), \qquad (1)$$

где (c_1, r_1) и (c_2, r_2) являются координатами точек (ячеек таблицы весов) до ближайшей точки (ячейки) изображения эталона с координатами (c_p, r_p) (рис. 2, *а*). В качестве выражений меры близости могут использоваться и другие более простые или сложные выражения [9], отличные от (1).

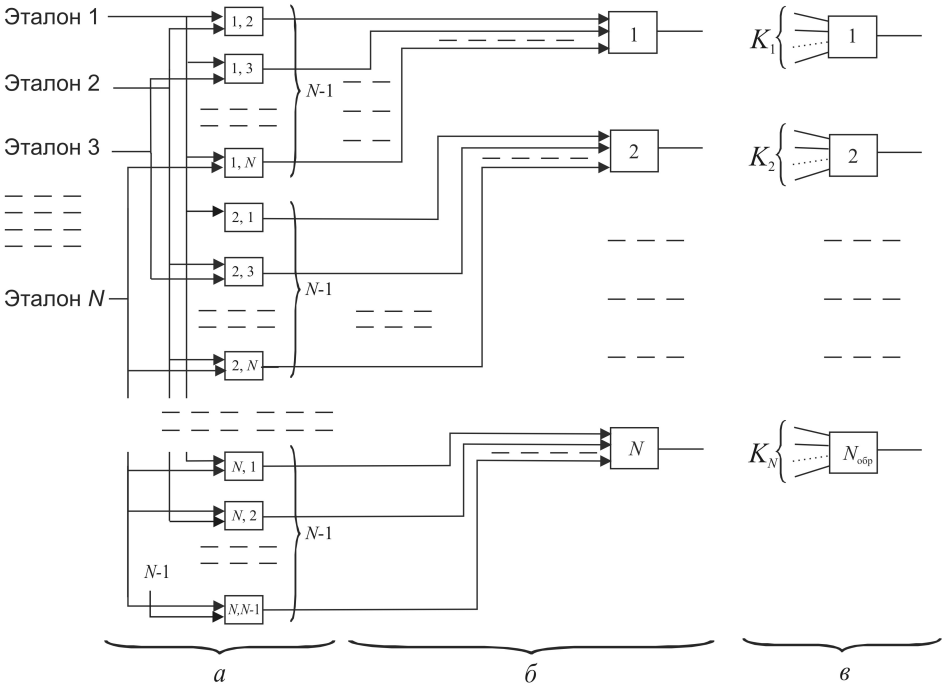

Рис. 1. Архитектура сети на основе метода ближайшего соседа для N эталонов и $N_{\text{обр}}$ образов (классов)

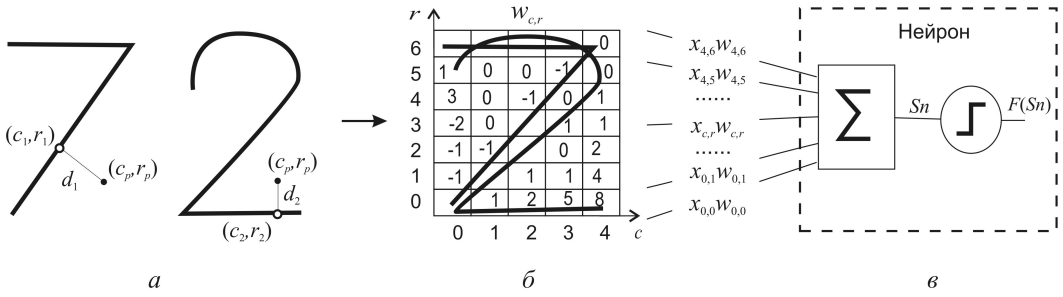

Рис. 2. Расстояния d_1 и d_2 для точки (c_p, r_p) ($а$); таблица весов для эталонов «2» и «7» ($б$); нейрон с пороговой функцией активации ($в$)

Функции состояния $Sn_{i,j}^{(1)}$ и активации $f\left(Sn_{i,j}^{(1)}\right)$ для каждого нейрона первого слоя определяются по формулам

$$Sn_{i,j}^{(1)} = \sum_{r=0}^{R}\sum_{c=0}^{C} x_{c,r} w_{c,r}^{(1)}, \quad f\left(Sn_{i,j}^{(1)}\right) = \begin{cases} 1, & \text{если } Sn_{i,j}^{(1)} < 0, \\ 0, & \text{если } Sn_{i,j}^{(1)} > 0, \end{cases}$$

где C, R — количество столбцов и строк в таблице весов (на рис. 2, $б$ $C = 5$, $R = 7$). Значения весов для каждого входа нейронов второго и третьего слоя равны 1:

$$w_{i,j}^{(2)} = w_{i,j}^{(3)} = 1.$$

Количество нейронов второго слоя равно количеству используемых эталонов $n_2 = N$. Значения функций состояния и активации для k-го нейрона второго слоя определяются по выражениям

$$Sn_k^{(2)} = \sum_{j=1,\, j\neq k}^{N} f\left(Sn_{k,j}^{(1)}\right); \tag{2}$$

$$f\left(Sn_k^{(2)}\right) = \begin{cases} 1, & \text{если } Sn_k^{(2)} \geqslant (N-1) = B^{(2)}, \\ 0, & \text{если } Sn_k^{(2)} < (N-1) = B^{(2)}. \end{cases} \tag{3}$$

Здесь $B^{(2)} = N-1$ — пороговое значение нейрона второго слоя. Если выход k-го нейрона второго слоя равен $y_k^{(2)} = 1$, то это означает, что объекту $\bar{x}$ на входе нейронной сети соответствует k-й эталон. Нейроны третьего слоя объединяют выходы эталонов одного k-го образа в единый выход $y_k^{(3)}$. Функции состояния $Sn_k^{(3)}$ и активации $f\left(Sn_k^{(3)}\right)$ для нейрона третьего слоя определяются по формулам

$$Sn_k^{(3)} = \sum_{i\in k}^{K} f\left(Sn_i^{(2)}\right), \quad f\left(Sn_k^{(3)}\right) = \begin{cases} 1, & \text{если } Sn_k^{(3)} > 0, \\ 0, & \text{если } Sn_k^{(3)} = 0. \end{cases}$$

Если на рис. 1 каждому распознаваемому образу соответствует один эталон, то архитектура нейронной сети преобразуется в двуслойную нейронную сеть (рис. 1, *а,б*), где количество нейронов второго слоя равно количеству распознаваемых образов $N_{\text{обр}}$.

1. Постановка задачи и методы решения

Как уже было сказано, архитектура нейронной сети на рис. 1 реализует метод ближайшего соседа, где значения весов первого слоя вычисляются аналитически по формулам (1) на основе выделенного набора эталонов. При этом напрашивается вопрос: может ли данная архитектура нейронной сети обучаться классическими алгоритмами обучения без использования выделенных эталонов и аналитических выражений?

Представим каждый нейрон на рис. 1, *а,б* первого слоя как отдельный блок нейронной сети ($NN_{i,j}$) рис. 3, *а*, выполняющий задачу разделения двух образов. При этом у каждого блока $NN_{i,j}$ имеется один бинарный выход $y_{i,j}^{(1)} \in \{0,1\}$, значение которого указывает на близость объекта $\bar{x}$ к образу i или j. Первый блок $NN_{1,2}$ на рис. 3, *а* выполняет разделение образов 1 и 2, второй блок $NN_{1,3}$ разделяет образы 1 и 3 и т. д., вплоть до блока $NN_{1,N-1}$, разделяющего образы 1 и $N-1$. Затем следуют блоки $NN_{i,j}$ для разделения пар образов $\{2,1\}$, $\{2,3\}$, …, $\{2,N-1\}$, $\{3,1\}$ и т. д. Каждый блок $NN_{i,j}$ обучается отдельно на разделение своей пары образов. Для обучающей выборки блока $NN_{i,j}$ используются только объекты образов i,j (рис. 3). Для обучения блока $NN_{i,j}$ могут использоваться любые известные классические алгоритмы обучения. Если на вход нейронной сети (рис. 3, *а*) подаётся объект $\bar{x}$, который принадлежит k-му образу, то на выходах первого слоя будут активны $N-1$ выходов, начиная от блока $NN_{k,j}$ до блока $NN_{k,N}$ ($k \neq j$), а на выходе второго слоя (рис. 3, *б*), согласно формулам (2) и (3), будет активен выход k-го нейрона $y_k^{(2)} = 1$. На выходах остальных нейронов второго слоя ($i \neq k$) будем иметь $y_i^{(2)} = 0$, поскольку для каждого $i \neq k$ существует один блок $NN_{i,k}$, у которого $y_{i,k}^{(1)} = 0$. Таким образом, максимальное значение $\max\left(Sn_{i\neq k}^{(2)}\right) = N-2$, тогда из (3) следует $f\left(Sn_{i\neq k}^{(2)}\right) = 0$.

В простейшем случае блок $NN_{i,j}$ может состоять из одного нейрона. Это будет в том случае, когда разделяемые образы достаточно удалены друг от друга и легко разделяются одной гиперплоскостью. В более сложных случаях блок может состоять из нескольких нейронов и слоёв. Можно также установить фиксированное значение нейронов и слоёв для каждого блока $NN_{i,j}$. Нейроны блока $NN_{i,j}$ могут быть

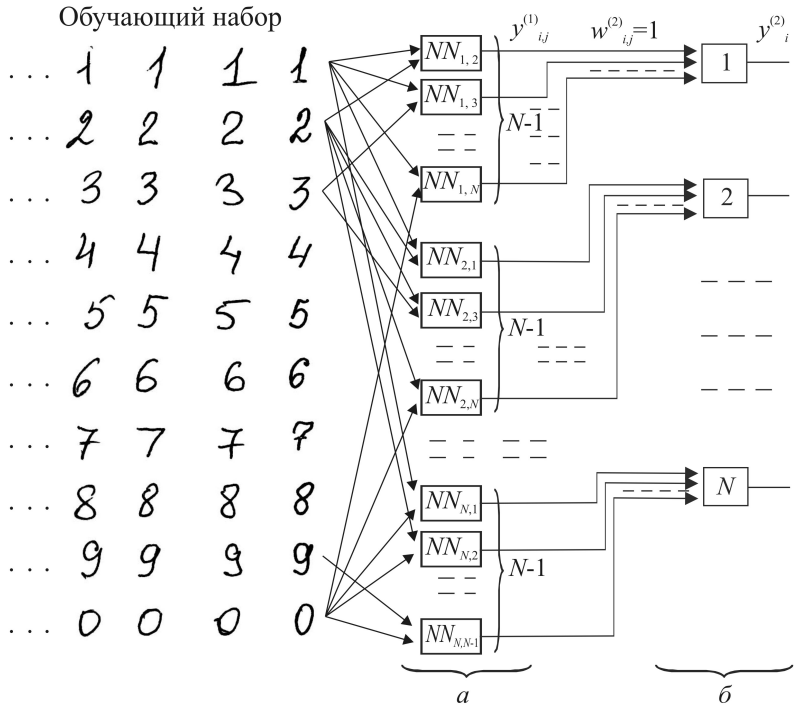

Рис. 3. Схема обучения нейронной сети

как с пороговыми, так и с непрерывными функциями активации. В последнем случае необходим дополнительный нейрон на выходе блока $NN_{i,j}$, преобразующий выходное непрерывное значение блока $NN_{i,j}$ в бинарное значение $y^{(1)}_{i,j} \in \{0,1\}$ (рис. 4). Для этого нейрона функция активации определяется по выражению

$$f\left(y_{\text{sigm}}\right) = \begin{cases} 1, & \text{если } y_{\text{sigm}} > 0{,}5, \\ 0, & \text{если } y_{\text{sigm}} \leqslant 0{,}5. \end{cases}$$

В качестве алгоритмов обучения для блоков $NN_{i,j}$ могут использоваться любые известные классические алгоритмы для нейронных сетей прямого распространения.

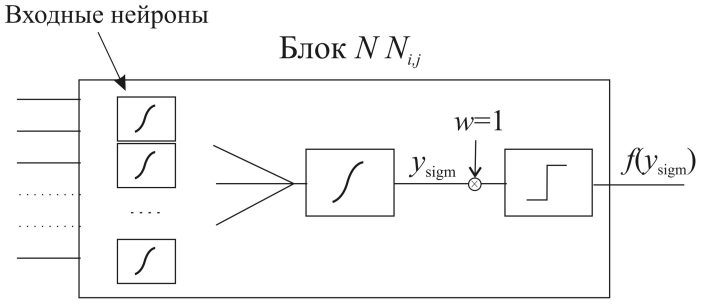

Рис. 4. Преобразование выхода $NN_{i,j}$ в бинарную форму

Недостатком архитектуры нейронной сети является большое количество n_1 блоков $NN_{i,j}$, которое для схемы на рис. 3 увеличивается по формуле $n_1 = (N-1)\,N$. Здесь возможно уменьшить это количество в 2 раза, если исключить блоки, повторяющие разделение одноимённых пар образов, например таких, как $\{1,2\}$ и $\{2,1\}$ [7]. В этом

случае выход $y_{i,j}^{(1)}$ блока $NN_{i,j}$, выполняющего разделение образов i и j, будет соединяться с i-м нейроном второго слоя ($y_{i,j}^{(1)} = x_i^{(2)}$), а инвертированное значение выхода $y_{i,j}^{(1)}$ — с j-м нейроном второго слоя ($\bar{y}_{i,j}^{(1)} = x_j^{(2)}$) (рис. 5). Можно предположить, что количество блоков $NN_{i,j}$ может быть ещё меньше, если исходить из того, что не все блоки NN являются жизненно важными и, возможно, некоторые из них могут быть удалены. Последнее утверждение требует дополнительных исследований.

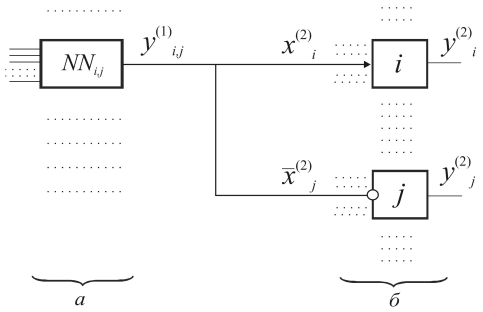

Рис. 5. Сжатие сети путём исключения блока NN_{ji}

Таким образом, обучение нейронной сети (рис. 3) сводится к независимому обучению каждого отдельного блока $NN_{i,j}$. Обучение нейронной сети становится проще, понятнее и надёжнее, поскольку обучение множества N образов приводится к задаче попарного разделения двух образов. Исходя из этого, понадобится меньшая обучающая выборка по сравнения с классическими нейронными сетями, разделяющими много образов. Количество эпох обучения по этой же причине будет также меньше. Нет также необходимости обучать слой, приведённый на рис. 3, *б*. Обучение каждого блока $NN_{i,j}$ выполняется независимо от других блоков, что позволяет легко расширять нейронную сеть, добавляя новые распознаваемые образы к общей задаче. При этом нет необходимости заново переобучать всю сеть, предыдущие весовые и пороговые значения сохраняются. Возможное количество распознаваемых образов потенциально может быть очень большим ($N \to \infty$). Для каждого последующего $(N+1)$-го образа к схеме нейронной сети на рис. 3 для каждого k-го образа добавляются по одному блоку $NN_{k,N+1}$ (всего N блоков) и N блоков $NN_{N+1,j}$ для $(N+1)$-го образа. Итого $2N$ блоков, где N — последнее значение количества образов. В случае сжатого варианта нейронной сети (рис. 5) для каждого нового образа добавляются только N новых блоков $NN_{i,j}$. При добавлении образа пороговое значение $B^{(2)}$ нейрона второго слоя (3) (рис. 3, *б*) увеличивается на 1.

Покажем работу схемы нейронной сети на рис. 3 и 5 на примере задачи распознавания образов трёх символов «A», «B», «C». Пусть для каждого образа существует обучающая выборка с n элементами. На рис. 6 приведена сжатая схема нейронной сети (рис. 5) для данной задачи распознавания. Поскольку количество блоков на схеме равнозначно количеству нейронов первого слоя на рис. 6, *б*, то, исходя из выражения

$$n^{(1)} = C_N^2 = \frac{N!}{2!\,(N-2)!} = \frac{N\,(N-1)}{2} = \frac{3\,(3-1)}{2} = 3,$$

определяющего количество нейронов первого слоя из общего количества распознаваемых образов на основе метрического метода распознавания [7], понадобится три блока $NN_{i,j}$, три нейрона второго слоя и три выхода нейронной сети Y_i. Если соответствие

выходов нейронной сети к распознаваемому образу обозначено так, как на рис. 6 (в порядке «A», «B», «C» сверху вниз), то, согласно схеме, первый блок $NN_{1,2}$ выполняет разделение образов «A», «B», блок $NN_{1,3}$ — образов «A», «C», блок $NN_{2,3}$ — «B», «C». На рис. 6, *а* каждому блоку соответствует бинарная матрица символа. При обучении на первую матрицу подаются бинарные изображения обучающих выборок образов «A», «B», на вторую — образов «A», «C» и на третью — образов «B», «C». При программной реализации данная матрица может быть в единственном числе и отдельно использоваться для каждого блока. Размерность бинарной матрицы (R — количество строк, C — столбцов) может быть различной, она определяет количество связей (RC) для каждого входного нейрона блока $NN_{i,j}$. Количество входных нейронов блока $NN_{i,j}$ может быть также различно. Чем больше размерность бинарной матрицы X и чем больше количество входных нейронов в одном блоке, тем лучше будет результат распознавания на выходе данного блока после его обучения. Каждый блок обучается отдельно при помощи известных классических алгоритмов обучения, то есть классическим способом определяется таблица весов $W_{ij}^{(1)}$ отдельно для каждого блока $NN_{i,j}$. Весовые значения для нейронов второго слоя равны $w_{ij}^{(2)} = 1$. Поскольку используется сжатая форма нейронной сети, на входах нейронов 2 и 3 второго слоя (рис. 6, *в*) используются инвертированные значения выходов блоков $y_{1,2}$, $y_{1,3}$, $y_{2,3}$, которые равнозначно заменяют соответствующие входы на расширенной форме схемы нейронной сети на рис. 1. Для схемы на рис. 6 на трёх выходах нейронной сети значение «100» соответствует образу символа «A», «010» — «B» и «001» — «C».

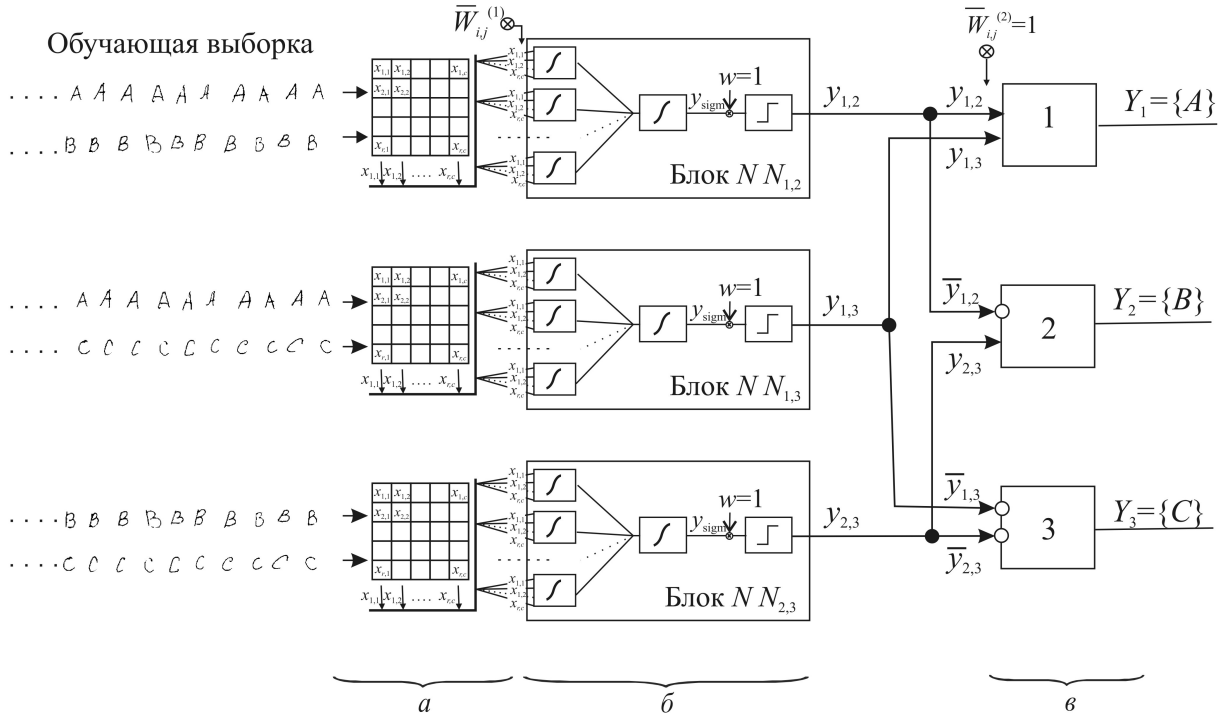

Рис. 6. Схема нейронной сети для задачи распознавания трёх образов: бинарные таблицы (*а*), слой блоков NN_{ij} (*б*), второй слой нейронной сети (*в*)

Предположим, что после проверки результатов обучения контрольной выборкой для каждого блока количество неверно идентифицированных символов составило для первого блока k_1 символов, для второго — k_2 и для третьего блока — k_3 символов. В табл. 1 и 2 приведены значения состояний и выходов нейронов второго слоя (рис. 6, *в*)

для распознаваемого символа «A» в зависимости от значений выходов блоков $NN_{1,2}$ и $NN_{1,3}$. Поскольку выход блока $NN_{1,2}$ определяет разделение образов «A», «B» ($y_{1,2} = 1$ для символа «A» и $y_{1,2} = 0$ для символа «B»), а выход блока $NN_{1,3}$ определяет разделение образов «A», «C» ($y_{1,3} = 1$ для символа «A» и $y_{1,3} = 0$ для символа «C»), из табл. 1 можно видеть, что для всех блоков с правильно идентифицированными символами «A» ($y_{1,2} = 1$ и $y_{1,3} = 1$) в соответствии с выражениями (2) и (3) на выходах нейронной сети формируется сигнал «100», соответствующий образу символа «A».

Таблица 1

Значения выходов всех блоков и нейронов второго слоя для распознаваемого символа «A» для случая, когда блоки $NN_{1,2}$ и $NN_{1,3}$ распознают символ «A» правильно

№ блоков	Выходы блоков	№ нейрона второго слоя	Значение нейрона второго слоя S_n	Выход нейрона второго слоя Y_i, $N-1=2$	Выбранный образ
1, 2	$y_{1,2} = 1$	1	$S_{n_1} = y_{1,2} + y_{1,3} = $ $= 1 + 1 = 2 = N - 1$	$Y_1 = 1$ $(= 2)$	«A»
1, 3	$y_{1,3} = 1$	2	$S_{n_2} = \overline{y_{1,2}} + y_{2,3} = $ $= 0 + (0 \text{ или } 1) = $ $= (0 \text{ или } 1) < N - 1$	$Y_2 = 0$ (< 2)	–
2, 3	$y_{2,3} = $ $= (1 \text{ или } 0)$	3	$S_{n_3} = \overline{y_{1,3}} + \overline{y_{2,3}} = $ $= 0 + (1 \text{ или } 0) = $ $= (1 \text{ или } 0) < N - 1$	$Y_3 = 0$ (< 2)	–

Таблица 2

Значения выходов всех блоков и нейронов второго слоя для распознаваемого символа «A» для случая, когда блок $NN_{1,2}$ распознаёт символ «A» неправильно

№ блоков	Выходы блоков	№ нейрона второго слоя	Значение нейрона второго слоя S_n	Выход нейрона второго слоя Y_i, $N-1=2$	Выбранный образ
1, 2	$y_{1,2} = 0$	1	$S_{n_1} = y_{1,2} + y_{1,3} = $ $= 0 + 1 = 1 < N - 1$	$Y_1 = 0$ (< 2)	–
1, 3	$y_{1,3} = 1$	2	$S_{n_2} = \overline{y_{1,2}} + y_{2,3} = $ $= 1 + (0 \text{ или } 1) = $ $= (1 \text{ или } 2) \leqslant N - 1$	$Y_2 = 0$ $(\leqslant 2)$	«B» или –
2, 3	$y_{2,3} = $ $= (1 \text{ или } 0)$	3	$S_{n_3} = \overline{y_{1,3}} + \overline{y_{2,3}} = $ $= 0 + (0 \text{ или } 1) = $ $= (0 \text{ или } 1) < N - 1$	$Y_3 = 0$ (< 2)	–

Напротив, для случая неправильной идентификации символа «A» в блоках $NN_{1,2}$ или $NN_{1,3}$ на выходе итоговой нейронной сети формируется сигнал «010» или «001», соответствующий символам «B» или «C» (табл. 2). Таким образом, можно видеть, что результат работы нейронной сети (рис. 6) зависит только от правильности работы блоков $NN_{i,j}$, то есть, иными словами, от того, насколько хорошо обучены блоки $NN_{i,j}$.

Если множество ошибочно идентифицируемых символов для каждого блока обозначить как $E(A,B)$, $E(A,C)$, $E(B,C)$ (рис. 7), то максимальная ошибка работы сети на контрольной выборке определяется как сумма ошибок всех блоков нейронной сети; для схемы на рис. 6 $K_{\max} = k_1 + k_2 + k_3$ в том случае, когда множества неправильно идентифицированных символов каждого образа не пересекаются друг с другом (рис. 7, *а*). Возможная минимальная ошибка может быть равна количеству элементов множества с максимальным количеством элементов, например $E(A,B)$ (рис. 7, *в*). Для рис. 7, *б* ошибка K определяется количеством элементов множества E, которое формируется по выражению

$$|E| = |E(A,B)| + |E(A,C)| - |E(A,B) \cap E(A,C)| + |E(B,C)| - |E(A,B) \cap E(B,C)|.$$

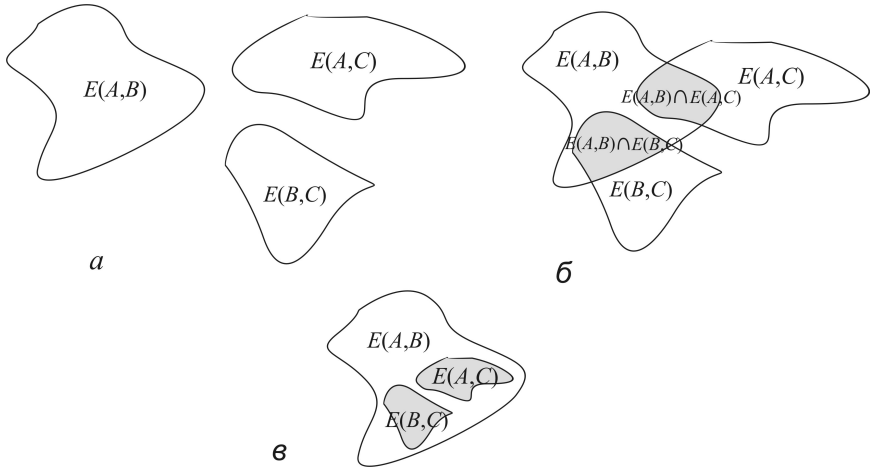

Рис. 7. Непересекающиеся множества ошибок для каждого блока (*а*); пересекающиеся множества (*б*) и (*в*)

Нужно также отметить, что, в отличие от схемы на рис. 1, в которой входные данные представляются в виде бинарной матрицы, для схем на рис. 3 и 6 значения входных матриц могут быть отличны от 0 и 1, поскольку попарное разделение двух образов выполняется классической схемой нейронной сети. Это позволяет в качестве входных данных в схемах на рис. 3 и 6 использовать чёрно-белые растровые изображения, у которых каждый пиксель может принимать значение в диапазоне $0 \div 256$; например, на рис. 8 показано изображение символа «8» со значениями пикселей в серых оттенках.

Матрица на рис. 8 может быть нормирована, например, в диапазон $(0,1)$. Могут быть также использованы отдельные палитры цветных изображений, представленные в той или иной цветовой схеме, например RGB, CMYK и т. д. Это может позволить в качестве входных данных нейронных сетей на рис. 3 и 6 использовать цветные изображения.

Само представление входных данных может быть отлично от непосредственного использования пикселей изображения. В качестве входных данных могут быть использованы предварительно определённые, выделенные или вычисленные данные изображения, которые более информативны для решаемой задачи. Например, в работе [10] приведены подходы представления входных данных для различных задач, в частности возможные способы представления данных для задачи распознавания лиц, а также для задачи распознавания символов с использованием инвариантных способов представления символов [11], позволяющие узнавать изображения символов с учётом возможных сдвигов, поворотов и масштабирования данного изображения.

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	4	62	146	182	254	254	181	176	139	15	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	34	186	253	217	208	136	136	136	166	232	99	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	61	242	208	111	3	0	0	0	0	0	18	32	107	43	0	0	0	0	0	0
0	0	0	0	0	0	0	0	156	242	23	0	0	0	0	0	0	0	13	191	181	6	0	0	0	0	0	0
0	0	0	0	0	0	0	0	121	255	98	3	0	0	0	0	0	8	194	225	12	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	169	253	120	3	0	0	0	0	128	247	51	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	3	111	244	169	19	0	14	131	249	117	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	59	241	235	72	142	229	66	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	25	218	254	231	36	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	133	253	221	33	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	19	237	111	196	217	19	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	174	138	0	23	193	204	18	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	96	224	0	0	0	25	218	169	3	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	215	138	0	0	0	0	86	253	99	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	215	97	0	0	0	0	3	162	214	11	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	215	97	0	0	0	0	0	118	253	68	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	185	157	0	0	0	0	0	40	254	98	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	50	244	61	0	0	0	0	112	244	58	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	174	251	142	59	83	167	244	111	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	6	133	253	253	253	169	61	3	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Рис. 8. Матрица символа «8» в диапазоне значений $0 \div 256$

Заключение

По сравнению с классическими и другими известными схемами нейронной сети, приведённая схема имеет ряд преимуществ, в частности:

1) архитектура сети сравнительно проста;
2) за исключением размерности матрицы и количества входных нейронов блока NN, все остальные параметры архитектуры сети строго определяются исходя из начального условия задачи — количества распознаваемых образов (классов);
3) количество распознаваемых образов может легко и последовательно увеличиваться и достигать очень больших значений, позволяя решать задачи с огромным количеством распознаваемых классов. При этом при добавлении новых образов (классов) предыдущие полученные весовые и пороговые значения не меняются;
4) обучение использует классические алгоритмы и при этом является более простым, поскольку сводится к разделению двух попарных образов. Следовательно, потребуется меньший объем обучающей выборки и количество эпох обучения;
5) простота попарного обучения также должна значительно уменьшить вероятность попадания нейронной сети в локальный минимум в процессе обучения.

Можно добавить, что рассмотренные возможности архитектуры нейронной сети на рис. 3 имеют некоторую схожесть с возможностями биологического мозга. В частности, в рассмотренной архитектуре объекты узнаются сразу целиком, как это обычно и

свойственно биологическому мозгу. Это отличает данную архитектуру, например, от архитектуры сверточных сетей [4 – 6], где объекты последовательно (послойно) анализируются от простейших деталей изображения до сложных. Кроме того, в предлагаемой архитектуре количество распознаваемых или добавляемых образов практически не ограничено, что также очень похоже на возможности биологического мозга, который может легко запоминать и узнавать огромное количество объектов и образов.

ЛИТЕРАТУРА

1. *Барский А. Б.* Нейронные сети: распознавание, управление, принятие решений. М.: Финансы и статистика, 2004.
2. *Головко В. Л.* Нейронные сети: обучение, организация и применение / под ред. А. И. Галушкина. Кн. 4. М.: ИПРЖР, 2001.
3. *Уоссермэн Ф.* Нейрокомпьютерная техника. Теория и практика. М.: Мир, 1992.
4. *LeCun Y., Bengio Y., and Hinton G.* Deep learning // Nature. 2015. No. 521. P. 436–444.
5. *Schmidhuber J.* Deep learning in neural networks: An overview // Neural Networks. 2015. No. 61. P. 85–117.
6. `http://www.neuralnetworksanddeeplearning.com/index.html` — Neural Networks and Deep Learning.
7. *Geidarov P. Sh.* Neural networks on the basis of the sample method // Automatic Control and Computer Sci. 2009. V. 43. No. 4. P. 203–210.
8. *Geidarov P. Sh.* Multitasking application of neural networks implementing metric methods of recognition // Autom. Remote Control. 2013. V. 74. No. 9. P. 1474–1485.
9. *Биргер И. А.* Техническая диагностика. М.: Машиностроение, 1978.
10. *Гейдаров П. Ш.* Алгоритм реализации метода ближайшего соседа в многослойном персептроне // Труды СПИИРАН. 2017. Вып. 51. С. 123–151.
11. *Васин Д. Ю., Аратский А. В.* Распознавание символов на основе инвариантных моментов графических изображений // 25 Междунар. конф. GraphiCon 2015. М., 2015. С. 259–264.

REFERENCES

1. *Barskiy A. B.* Neyronnye seti: raspoznavanie, upravlenie, prinyatie resheniy [Neural Networks: Recognition, Management, Decision Making]. Moscow, Finansy i Statistika Publ., 2004. (in Russian)
2. *Golovko V. L.* Nejronnye seti: obuchenie, organizacija i primenenie [Neural Networks: Training, Organization and Application]. Book 4, Moscow, IPRJR Publ., 2001. (in Russian)
3. *Wasserman F.* Neyrokomp'yuternaya tekhnika. Teoriya i praktika [Neurocomputing. Theory and Practice]. Moscow, Mir Publ., 1992. (in Russian)
4. *LeCun Y., Bengio Y., and Hinton G.* Deep learning. Nature, 2015, no. 521, pp. 436–444.
5. *Schmidhuber J.* Deep learning in neural networks: An overview. Neural Networks, 2015, no. 61, pp. 85–117.
6. `http://www.neuralnetworksanddeeplearning.com/index.html` — Neural Networks and Deep Learning.
7. *Geidarov P. Sh.* Neural networks on the basis of the sample method. Automatic Control and Computer Sci., 2009, vol. 43, no. 4, pp. 203–210.
8. *Geidarov P. Sh.* Multitasking application of neural networks implementing metric methods of recognition. Autom. Remote Control, 2013, vol. 74, no. 9, pp. 1474–1485.
9. *Birger I. A.* Tekhnicheskaya diagnostika [Technical Diagnostics]. Moscow, Mashinostroenie Publ., 1978. (in Russian)

10. *Geidarov P. Sh.* Algoritm realizacii metoda blizhajshego soseda v mnogoslojnom perseptrone [Algorithm for realizing the method of the nearest neighbor in a multilayer perceptron]. Trudy SPIIRAN, 2017, no. 51, pp. 123–151. (in Russian)
11. *Vasin D. Yu. and Aratskiy A. V.* Raspoznavanie simvolov na osnove invariantnyh momentov graficheskih izobrazhenij [Character recognition based on invariant moments of graphic images]. 25th Intern. GraphiCon Conf., Moscow, 2015, pp. 259–264. (in Russian)

УДК 519.852

МЕТОД ПОСЛЕДОВАТЕЛЬНОЙ АКТИВАЦИИ ОГРАНИЧЕНИЙ В ЛИНЕЙНОМ ПРОГРАММИРОВАНИИ

В. С. Колосов

г. Москва, Россия

Представлен алгоритм решения задачи линейного программирования посредством процедуры последовательной активации ограничений (включения в расчёт одного за другим) с удержанием состояния оптимальности на сгенерированной последовательности вложенных многогранников. Вследствие сжатия области допустимых решений при критерии оптимальности «max» целевая функция на каждом шаге убывает (движение к максимуму сверху), в противоположность росту в других методах (снизу). Компьютерные эксперименты демонстрируют преимущества программной реализации этого алгоритма перед опцией симплекс-метода программы `linprog` библиотеки MATLAB в скорости и полноте выводимой информации.

Ключевые слова: *линейное программирование, активация ограничения, MATLAB, компьютерный эксперимент.*

DOI 10.17223/20710410/41/11

METHOD FOR SEQUENTIAL ACTIVATION OF LIMITATIONS IN LINEAR PROGRAMMING

V. S. Kolosov

Moscow, Russia

E-mail: vs.kolosov@yandex.ru

In this paper, we present an algorithm for solving a problem in linear programming by means of procedure of sequential activation of limitations (inclusions in calculation are made one after another) with retention of the optimality status in the generated sequence of nested polyhedrons. Due to compression of the admissible solutions area, the objective function decreases on each step at the “max” criterion of an optimality (the movement to a maximum “from above”) contrary to its growth in other methods (“from below”). Geometrically, the motion to the maximum starts from the trivially determined starting point and continues along the broken line outside of the polytope of admissible solutions. In the theoretical justification of the algorithm, the signs for the incompatibility of the system of conditions in the problem, for the uniqueness and nonuniqueness of the solution, and for its unboundedness are formulated and proved. Computer experiments demonstrate the advantages of the program implementation of this algorithm in speed and completeness of the output information over the simplex-method option of the MATLAB’s library linprog program.

Keywords: *linear programming, activation of limitation, MATLAB, computer experiment.*

Введение

Число существенно различных методов условной линейной оптимизации невелико. Еще в 1939 г. Л. В. Канторович сформулировал ряд задач линейного программирования и для их решения предложил метод разрешающих множителей [1]. В 1947 г. Дж. Данцигом был создан симплекс-метод, нашедший отражение в отечественной литературе в изложении автора в переводе [2]. Разработанные в то время методы последовательного улучшения плана, сокращения оценок и сокращения невязок классифицируются по типу решаемых ими задач — прямой, двойственной или сразу обеих [3]. В середине 1960-х годов И. И. Дикин в [4] предложил решение задачи линейного программирования методом внутренних точек. Поиск новых подходов к оптимизации дал эффектный, но вычислительно неэффективный метод эллипсоидов Л. Г. Хачияна [5], принципиально важным достоинством которого явилось выражение времени счёта полиномом от размерности задачи. Полиномиальность варианта метода внутренних точек Н. Кармаркара от 1984 г. в совокупности с его вычислительной эффективностью послужила причиной его патентования в США. Следствием изысканий в данном направлении было множество публикаций, в их числе [6 – 8]. Метод внутренних точек глубоко проработан и как эффективный оптимизатор задействован в ряде программных продуктов, в том числе в MATLAB.

К недавним разработкам относятся метод оператора-проектора [9] и скелетный алгоритм [10]. Суть первого заключается в проектировании произвольной точки на гиперплоскости ограничений вплоть до попадания в область допустимых решений и удержании последующих перемещений в заданных границах согласно критерию оптимальности. Скелетный алгоритм не требует обращения матриц и подменяет исходную задачу рядом вспомогательных задач пониженной размерности. В процессе тестовых расчётов установлено быстрое продвижение к оптимуму в начале вычислительного процесса и его резкое замедление по мере приближения к цели. Метод внутренних точек является приближённым и находит решение со сколь угодно малой заданной погрешностью, но полиномиален; ни один из известных точных алгоритмов таким важным свойством не обладает, за некоторыми исключениями. Так, при условии неотрицательности всех параметров задачи и совместности системы ограничений метод экспоненциальной аппроксимации [11] обещает находить точный оптимум в задаче с n переменными и $m \leqslant n$ равенствами всего за $(n - m)$ шагов. Алгоритм выявляет и исключает переменные с гарантированно нулевыми оптимальными значениями, и решение исходной задачи завершается решением системы уравнений с квадратной матрицей m-го порядка. Перспективы ослабления жёстких требований этого метода к исходным данным не просматриваются.

Перечисленные выше методы объединяют два общих свойства:

— многогранник допустимых решений подразумевается существующим в своем целостном виде, т. е. предполагается одновременная действенность всех ограничений системы условий;

— на каждом шаге поиск следующей точки производится с умыслом улучшения значения целевой функции в пределах области допустимых решений.

Недавние публикации свидетельствуют, что, несмотря на снижение интереса к проблематике линейного программирования, поиск новых путей в этой сфере продолжается. В [12] автор завершил разработку точного алгоритма решения задачи линейного программирования с использованием параметризации лимитов системы условий, запрограммировал его на языке Fortran-4 и успешно апробировал на ЕС-ЭВМ. С учётом

современных возможностей программных и аппаратных средств данный метод им основательно доработан и реализован средствами MATLAB.

Далее изложена доказательная база этого по сути геометрического метода решения общей задачи линейного программирования, отличительная черта которого состоит в следующем. Многогранник допустимых решений не существует изначально, а формируется путём последовательного включения в расчёт (активации с помощью параметризации лимитов) ограничений с удержанием оптимальности целевой функции на генерируемой последовательности вложенных многогранников. Начальным является гиперпараллелограмм, задаваемый ограничениями на переменные с возможно неограниченными лимитами, на котором оптимум определяется элементарно. Затем в его систему условий добавляется первое нетривиальное ограничение, быстрым пересчётом находится оптимум на построенном многограннике (вложенном в предыдущий) и т. д. Активируя неравенства, алгоритм последовательно усекает образующиеся многогранники до окончательной конфигурации, постепенно сужая область допустимых решений, вычисляя и фиксируя на каждой итерации алгебраические и геометрические характеристики текущего оптимума для последующих применений.

Метод оперирует с неравенствами и, исключая переменные через посредство равенств, понижает размерность задачи, но может работать напрямую и с равенствами, легко разбираясь с ситуациями вырожденности и маскировки равенств совокупностью неравенств. В результате сжатия области допустимых решений в задачах с критерием оптимальности «max» целевая функция на каждом шаге убывает или, по крайней мере, не возрастает (к максимуму сверху), в противоположность её традиционному увеличению в других методах (снизу). Геометрически происходит перемещение в оптимум вдоль некоторой ломаной прямой из тривиально определяемой точки, которая для многогранника допустимых решений может оказаться как граничной, так и внешней. В процессе активации ограничений выявляется несовместность системы условий с указанием виновных в этом ограничений, вычисляются координаты оптимальной вершины и смежных с ней в случае неединственности, а также определяются двойственные переменные как оценки чувствительности.

Разработанный алгоритм реализован автором в виде программы `mpao` в среде MATLAB и апробирован на обширном экспериментальном материале. Показана простота процедуры ввода исходных данных и запуска программы. Установлено преимущество по быстродействию и широте выводимой информации программы `mpao` по сравнению с опцией симплекс-метода программы `linprog` библиотеки MATLAB на примерах задач с размерностями до 1000×1000.

1. Теоретическое обоснование алгоритма

В евклидовом пространстве E^n решается задача линейного программирования в самой общей постановке с разделением ограничений на «многофакторные», содержащие несколько переменных, и «координатные» (однофакторные), устанавливающие диапазоны возможных изменений собственно переменных при допущении отрицательности и неограниченности всех фигурирующих здесь величин включая сами переменные:

$$L(x) = cx \to \max, \quad c, x \in E^n,$$

$$\begin{cases} -\infty \leqslant \underline{b}^{(i)} \leqslant a_i x \leqslant \overline{b}^{(i)} \leqslant +\infty, a_i = \left(a_i^{(1)} \dots a_i^{(n)}\right), i = 1, \dots, m \text{ (многофакторные)}, \\ -\infty \leqslant \underline{d}^{(j)} \leqslant x^{(j)} \leqslant \overline{d}^{(j)} \leqslant +\infty, \quad j = 1, \dots, n \text{ (координатные)}. \end{cases} \tag{1}$$

Здесь сам вектор нумеруется нижним индексом, а его координаты — верхним в скобках. Многофакторные и координатные ограничения, высекая в E^n многогранник до-

пустимых решений (МДР) M, геометрически равнозначны. Первым отвечают гиперплоскости с задаваемой нормалями a_i ориентацией в пространстве, а вторым — ортогональные осям координат. Равенства реализуются уравниванием значений нижней и верхней границ разрешённых изменений и в отдельную группу не выделяются.

Равенствам отвечает размещение МДР в пересечении соответствующих гиперплоскостей, что открывает принципиальную возможность понижения размерности задачи за счёт фактического исключения (в отличие от симплекс-метода) части переменных. Это обстоятельство порождает две алгоритмические ветви. Следуя первой ветви, из равенства $a_i^{(1)}x^{(1)} + \ldots + a_i^{(n)}x^{(n)} = b^{(i)}$ выражается переменная с ненулевым коэффициентом; пусть это $x^{(n)} = \alpha_i^{(1)}x^{(1)} + \ldots + \alpha_i^{(n-1)}x^{(n-1)} + \alpha_i^{(n)}$. При её исключении из остальных многофакторных ограничений и целевой функции координатное ограничение $\underline{d}^{(n)} \leqslant x^{(n)} \leqslant \overline{d}^{(n)}$ преобразуется в многофакторное неравенство $\beta_i^{(n)} \leqslant \alpha_i^{(1)}x^{(1)} + \ldots + \alpha_i^{(n-1)}x^{(n-1)} \leqslant \beta_i^{(n)}$ с меньшим числом переменных. В итоге размерность пространства решений понижается, но только за счет переменных при сохранении количества многофакторных ограничений. Исключением равенств из системы условий заканчивается подготовительный шаг алгоритма. С этого момента МДР считается полноразмерным ($\dim M = n$), т. е. задаётся только неравенствами. Особенности другой ветви алгоритма без исключения равенств из системы условий будут разъяснены позже.

Совокупности неравенств в E^n отвечает выпуклый многогранник M. Её любая подсистема ранга n представляет собой неограниченный многогранник $\dot{M}(\dot{x}) \supseteq M$ с единственной вершиной в точке $\dot{x}$, являющейся решением соответствующей системы линейных равенств; пусть это $\{a_i x = b^{(i)} : i = 1, \ldots, n\}$. В параметрическом представлении

$$\dot{M}(\dot{x}) = \Big\{x = \dot{x} + \sum_{j=1}^{n} \beta_j v_j(\dot{x}),\ \ \beta_j \geqslant 0,\ \ j = 1, \ldots, n\Big\}, \tag{2}$$

где $v_j(\dot{x})$ — направляющие векторы рёбер $R_j(\dot{x}) = \{x = \dot{x} + tv_j(\dot{x}),\ t \geqslant 0\}$ многогранника $\dot{M}$, выходящих из вершины $\dot{x}$ и образованных пересечением $n-1$ гиперплоскостей $G_i = \{x : a_i x = b^{(i)}\}$. Тогда $R_j(\dot{x}) = \bigcap\limits_{i=1, i\neq j}^{n} G_i$ и $\{\dot{x}\} = R_j(\dot{x}) \cap G_j$. Алгебраически $R_j(\dot{x}) = \{x : a_i x = b^{(i)},\ i = 1, \ldots, n,\ i \neq j;\ a_j x \leqslant b^{(j)}\}$, где оно индексируется номером того единственного ограничения, которое в его образовании не участвует. Возьмём произвольную точку $x_j \in R_j(\dot{x})$ и построим вектор $v_j(\dot{x}) = x_j - \dot{x}$, который, будучи направляющим вектором, удовлетворяет однородной системе ранга $n-1$:

$$\{a_i v_j(\dot{x}) = a_i x_j - a_i \dot{x} = b^{(i)} - b^{(i)} = 0,\ \ i = 1, \ldots, j-1, j+1, \ldots, n.$$

Этот вектор ортогонален нормалям $n-1$ гиперплоскостей и может быть найден с помощью алгоритма Грама — Шмидта, который линейно независимую совокупность векторов $u_1, \ldots, u_{n-1} \in E^n$ рекуррентными соотношениями ортогонального проектирования

$$u_i^0 = u_i,\ u_i^k = u_i^{k-1} - \left(u_i^{k-1}\tilde{u}_k^{k-1}\right)\tilde{u}_k^{k-1},\ \tilde{u}_i^{i-1} = u_i^{i-1} / \parallel u_i^{i-1} \parallel,\ k < i < n \tag{3}$$

(k — «порядок» проекции) преобразует в ортонормированную систему $\{\tilde{u}_i^{i-1},\ i = 1, \ldots, n-1 :\ \tilde{u}_i^{i-1}\tilde{u}_j^{j-1} = \delta_j^i\}$, δ_j^i — символ Кронекера. Очевидно, что при линейной независимости $\{u_0, u_1, \ldots, u_{n-1}\}$ решением системы $\{u_i v = 0 : i = 1, \ldots, n-1\}$ будет вектор $v(u_0, u_1, \ldots, u_{n-1}) = -u_0 + \sum\limits_{i=1}^{n-1}\left(u_0\tilde{u}_i^{i-1}\right)\tilde{u}_i^{i-1}$, где векторы $\tilde{u}_i^{i-1}$ вычислены проектором Грама — Шмидта (3). Легко проверяются такие свойства построенного вектора, как

$u_0 v < 0$ и $u_i v = 0$, $i \in \{1, \ldots, n-1\}$. Поскольку вершина $\dot{x}$ образована пересечением n гиперплоскостей, то, принимая поочередно в качестве u_0 их нормали, можно найти все направляющие векторы. В такой схеме для каждого вектора строится своя матрица ортогонализации. Поэтому более эффективно работает приём с одноразовым обращением матрицы координат нормалей и принятием в качестве направляющих векторов её столбцов, взятых с обратным знаком. С ростом размерности задачи метод ортогонализации быстро уступает обращению матрицы, которое представляет собой механизм вычисления направляющих векторов, выходящих из вершины многогранника.

Рассматриваемый метод опирается на общеизвестный факт достижения линейной функцией экстремума на выпуклом замкнутом многограннике по крайней мере в одной из его вершин. Метод реализует идею формирования из имеющихся ограничений начального многогранника M_0, с гарантией содержащего оптимум задачи x^*, нахождения на нём оптимальной вершины x_0^* и построения последовательности оптимальных вершин $x_0^* \to x_1^* \to \ldots \to x_m^* = x^*$ на промежуточных вложенных многогранниках $M_0 \supseteq M_1 \supseteq \ldots \supseteq M_m = M$. В этой цепочке текущий M_k получен усечением предыдущего M_{k-1} следующим k-м ограничением.

В качестве M_0 разумно принять ортогональный параллелепипед, образованный координатными ограничениями. В случае неопределённости какой-либо границы её значение сообразно конкретной ситуации полагается равным $\pm\infty$. Если $M \neq \varnothing$, то, безусловно, $x^* \in M_0$. В силу структурной примитивности M_0 максимум на нём линейной функции L достигается в вершине $x_0^* = \left(x_0^{(1)^*}, \ldots, x_0^{(n)^*}\right)$, координаты которой определяются простым правилом: $x_0^{(j)^*} = \begin{cases} \overline{d}^{(j)}, & \text{если } c^{(j)} \geqslant 0, \\ \underline{d}^{(j)}, & \text{если } c^{(j)} < 0, \end{cases} \quad j = 1, \ldots, n.$

Определение 1. Полупространство (неравенство) $P = \{x : ax \leqslant b\}$ назовём значимым для точки $\tilde{x} \in E^n$, если $\tilde{x} \notin P$ $(a\tilde{x} > b)$, и незначимым в противном случае $(a\tilde{x} \leqslant b)$. Значимость и незначимость гиперплоскости (равенства) $G = \{x : ax = b\}$ определяются через неравенство $a\tilde{x} \neq b$ и равенство $a\tilde{x} = b$.

В результате усечения M_0 первым многофакторным двусторонним неравенством образуется многогранник $M_1 = M_0 \cap \underline{P}_1 \cap \overline{P}_1$, где $\underline{P}_1, \overline{P}_1$ — встречные полупространства с непустым пересечением: $\underline{P}_1 = \left\{x : \underline{b}^{(1)} \leqslant a_1 x\right\}$, $\overline{P}_1 = \left\{x : a_1 x \leqslant \overline{b}^{(1)}\right\}$. Поскольку $M_1 \subseteq M_0$, то $\max\limits_{x \in M_1} L(x) \leqslant \max\limits_{x \in M_0} L(x)$. Характер последующих действий определяется значимостью данного ограничения для x_0^*, т. е. расположением величины $a_1 x_0^*$ относительно отрезка $\left[\underline{b}^{(1)}, \overline{b}^{(1)}\right]$. Неравенство $\underline{b}^{(1)} \leqslant a_1 x_0^* \leqslant \overline{b}^{(1)}$ означает, что $x_0^* \in M_1$, и потому максимум L на M_1 достигается в той же самой вершине, т. е. $x_1^* = x_0^*$. Таким образом, данное ограничение для x_0^* незначимо и можно перейти к построению многогранника M_2, усекая M_1 следующим ограничением. В противном случае проверяется $M_1 \neq \varnothing$, вычисляется x_1^* и т. д.

Рассмотрим k-й шаг алгоритма, к началу которого уже построен многогранник $M_{k-1} = M_0 \cap \underline{P}_1 \cap \overline{P}_1 \cap \ldots \cap \underline{P}_{k-1} \cap \overline{P}_{k-1}$; пусть $M_{k-1} \neq \varnothing$. Если $x_{k-1}^* \in M_k$, то $x_k^* = x_{k-1}^*$ и текущий шаг на этом завершается. Остаётся разобраться с ситуацией $x_{k-1}^* \notin M_k$. Здесь двустороннее неравенство $\underline{b}^{(k)} \leqslant a_k x \leqslant \overline{b}^{(k)}$ для вершины x_{k-1}^* может быть не выполнено лишь с одной стороны и либо $a_k x_{k-1}^* < \underline{b}^{(k)}$, либо $a_k x_{k-1}^* > \overline{b}^{(k)}$. Принципиального различия между типами этих двух неравенств нет, так как один сводится к другому умножением неравенства на -1.

В случае значимости ограничения $a_k x^*_{k-1} \notin \left[\underline{b}^{(k)}, \overline{b}^{(k)}\right]$ производятся основные расчёты. Пусть для определённости $x^*_{k-1} \notin \overline{P}_k$, граница которого $\overline{G}_k = \left\{x : a_k x = \overline{b}^{(k)}\right\}$. Параметризуем статическое полупространство $\overline{P}_k = \left\{x : a_k x \leqslant \overline{b}^{(k)}\right\}$, образовав динамическую структуру $\tilde{P}_k(\lambda) = \{x : a_k x \leqslant \lambda\}$, $\tilde{P}_k\left(\overline{b}^{(k)}\right) = \overline{P}_k$, $\tilde{G}_k(\lambda) = \{x : a_k x = \lambda\}$, $\tilde{G}_k\left(\overline{b}^{(k)}\right) = \overline{G}_k$.

Определение 2. Активацией значимого ограничения $a_k x \leqslant \overline{b}^{(k)}$ называется изменение параметра λ от $+\infty$ до $\overline{b}^{(k)}$, в результате чего движение гиперплоскости $\tilde{G}_k(\lambda)$ влечёт трансформацию $x^*_{k-1} \to x^*_k$.

Активация ограничения — динамическая процедура его приведения в активное состояние в случае значимости. Под активностью ограничения $P = \{x : ax \leqslant b\}$ в точке $\dot{x}$ традиционно понимается её принадлежность граничной гиперплоскости, т. е. $\dot{x} \in G = \{x : ax = b\}$ или $a\dot{x} = b$. Вычисление x^*_0 составляет содержание «нулевого» шага алгоритма, на котором активированы только все координатные ограничения.

Обозначим: $\{\dot{x}_j : j = 1, \ldots, r\}$ — концевые точки рёбер $R_j(x^*_{k-1})$ с направляющими векторами $v_j(x^*_{k-1})$, являющиеся вершинами M_{k-1}, смежными с x^*_{k-1}; $\tilde{P}_k(\lambda) = \{x : a_k x \leqslant \lambda\}$ — активируемое динамическое полупространство; $\lambda^* = a_k x^*_{k-1}$; $\dot{\lambda}_j = a_k \dot{x}_j$; $\tilde{M}_k(\lambda) = M_{k-1} \cap \tilde{P}_k(\lambda)$.

Сигналом к активации многофакторного ограничения $\overline{P}_k$ является его значимость, и на k-м шаге этому отвечает решение задачи оптимизации $L(\tilde{x}^*_k(\lambda)) = \max\limits_{\tilde{x} \in \tilde{M}_k(\lambda)} L(x)$, при том что $\tilde{M}_k(\lambda) = M_{k-1} \cap \tilde{P}_k(\lambda)$, $+\infty = \lambda \to \overline{b}^{(k)}$. По построению $\tilde{M}_k(\lambda)$ при всех λ из указанного диапазона относительно M_{k-1} не расширяется и $\tilde{M}_k(\lambda) \subseteq M_{k-1}$, $\tilde{M}_k(+\infty) = M_{k-1}$, $\tilde{M}_k\left(\overline{b}^{(k)}\right) = M_k$, $\tilde{x}^*_k\left(\overline{b}^{(k)}\right) = x^*_k$. Полупространство $\tilde{P}_k(\lambda)$ при $\lambda \geqslant \lambda^* = a_k x^*_{k-1}$ для x^*_{k-1} является незначимым. Его граница $\tilde{G}_k(\lambda)$ при $\lambda = \lambda^*$ проходит через x^*_{k-1}, а при $\lambda < \lambda^*$ отвечает за трансформацию $x^*_{k-1} \to \tilde{x}^*_k(\lambda)$.

Теорема 1. Если $a_k v_j(x^*_{k-1}) \geqslant 0$ для всех $j \in \{1, \ldots, r\}$, то для всех $\lambda < \lambda^*$ система условий задачи (1) несовместна ($M = \varnothing$).

Доказательство. Для любого $\lambda \leqslant \lambda^*$ по построению $\tilde{M}_k(\lambda) \subseteq M_{k-1}$, и в соответствии с представлением (2) для любого $\tilde{x} \in \tilde{M}_k(\lambda)$ с учётом $\beta_j \geqslant 0$ для всех $j \in \{1, \ldots, r\}$ находим $a_k \tilde{x} = a_k x^*_{k-1} + \sum\limits_{j=1}^{r} \beta_j a_k v_j(x^*_{k-1}) \geqslant a_k x^*_{k-1} = \lambda^*$. Тогда $\tilde{x} \notin \tilde{M}_k(\lambda)$ при $\lambda < \lambda^*$, т. е. $\tilde{M}_k(\lambda) = \varnothing$, а вследствие $M \subseteq \tilde{M}_k(\lambda)$ и $M = \varnothing$. ■

Замечание 1. Здесь сформирован механизм выявления противоречивости системы условий задачи в процессе активации ограничений.

Теорема 2. Пусть $\dot{x}$ — вершина выпуклого многогранника $M \neq \varnothing$ с направляющими векторами всех выходящих из неё рёбер $v_1(\dot{x}), \ldots, v_r(\dot{x})$. Максимум $L = cx$ на M достигается в $\dot{x}$, если $cv_j(\dot{x}) \leqslant 0$ для всех j; эта вершина единственна в случае $cv_j(\dot{x}) < 0$ для всех j. Верно и обратное.

Доказательство. Достаточность следует из (2). Из включения $M \subseteq \dot{M}(\dot{x})$ получаем, что $cx = c\dot{x} + \sum\limits_{j=1}^{r} \beta_j c v_j(\dot{x})$ для всех $x \in M$. Учитывая существование хотя бы одного $\beta_k > 0$ при $x \neq \dot{x}$ в случае $cv_k(\dot{x}) \leqslant 0$, получаем $cx \leqslant c\dot{x}$, а при строгом неравенстве $cv_k(\dot{x}) < 0$ таким же строгим будет неравенство $cx < c\dot{x}$.

Необходимость докажем от противного. Пусть $c\dot{x} > cx$ для всех $x \in M$, но существует $v_k(\dot{x})$, для которого $cv_k(\dot{x}) > 0$. Так как $x(t_0) = \dot{x} + t_0 v_k(\dot{x}) \in M$ при некотором малом $t_0 > 0$, то $cx(t_0) = c\dot{x} + t_0 cv_k(\dot{x}) > c\dot{x}$, что противоречит оптимальности $\dot{x}$. Далее, предположив единственность оптимума и существование $v_k(\dot{x})$, для которого $cv_k(\dot{x}) = 0$, получаем противоречие со сделанным предположением, поскольку $cx(t_0) = c\dot{x}$ при $x(t_0) \neq \dot{x}$. ■

Замечание 2. Так фиксируется неединственность решения.

Теорема 3. Если $M_k \neq \varnothing$, то для любого $\lambda \in \{\lambda^* - \varepsilon, \lambda^*\}$, где $\lambda^* = a_k x^*_{k-1}$, при достаточно малом $\varepsilon > 0$ имеет место $a_k \tilde{x}^*_k(\lambda) = \lambda$, чему геометрически соответствует $\tilde{x}^*_k(\lambda) \in \tilde{G}_k(\lambda)$.

Доказательство. Поскольку $\tilde{x}^*_k(\lambda) \in \tilde{M}_k(\lambda)$, то $a_k \tilde{x}^*_k(\lambda) \leqslant \lambda$. При допущении противного ($\tilde{x}^*_k(\lambda) \notin \tilde{G}_k(\lambda)$) это неравенство может быть только строгим ($a_k \tilde{x}^*_k(\lambda) < \lambda$). Из $\tilde{M}_k(\lambda) \subseteq M_{k-1}$ следует $\tilde{x}^*_k(\lambda) \in M_{k-1}$ и в силу выпуклости $M_{k-1} \supset [\tilde{x}^*_k(\lambda), x^*_{k-1}]$. По условию $\lambda < \lambda^*$ и, следовательно, концевые точки отрезка $\tilde{x}^*_k(\lambda)$ и x^*_{k-1} лежат строго по разные стороны гиперплоскости $\tilde{G}_k(\lambda)$, а значит, существует $x_\lambda \in (\tilde{x}^*_k(\lambda), x^*_{k-1})$, что $x_\lambda \in \tilde{G}_k(\lambda)$. В параметрическом представлении $(\tilde{x}^*_k(\lambda), x^*_{k-1}) = \{x(t) = \tilde{x}^*_k(\lambda) + t(x^*_{k-1} - \tilde{x}^*_k(\lambda)) : t \in (0,1)\}$; тогда существует $t_\lambda \in (0,1)$, при котором $x_\lambda = x(t_\lambda) = \tilde{x}^*_k(\lambda) + t_\lambda(x^*_{k-1} - \tilde{x}^*_k(\lambda))$. Отсюда находим $x^*_{k-1} = \tilde{x}^*_k(\lambda) + t_0(x_\lambda - \tilde{x}^*_k(\lambda))$, $t_0 = 1/t_\lambda > 1$. Сделанное допущение $\tilde{x}^*_k(\lambda) \notin \tilde{G}_k(\lambda)$ влечёт строгое неравенство $c\tilde{x}^*_k(\lambda) > cx$ для всех $x \in \tilde{G}_k(\lambda)$, в том числе $c\tilde{x}^*_k(\lambda) > cx_\lambda$ или $c(x_\lambda - \tilde{x}^*_k(\lambda)) < 0$. Посчитав $cx^*_{k-1} = c(\tilde{x}^*_k(\lambda) + t_0(x_\lambda - \tilde{x}^*_k(\lambda))) = c\tilde{x}^*_k(\lambda) + t_0 c(x_\lambda - \tilde{x}^*_k(\lambda)) < c\tilde{x}^*_k(\lambda)$, приходим к противоречию с оптимальностью x^*_{k-1} на $M_{k-1} \ni \tilde{x}^*_k(\lambda)$. ■

Замечание 3. Таким образом, если $M_k \neq \varnothing$, то либо сохраняется оптимальность предыдущей вершины и $x^*_k = x^*_{k-1}$, либо оптимальная вершина удовлетворяет активируемому ограничению как равенству с одной из его границ, что означает её принадлежность гиперплоскости $\underline{G}_k = \left\{x : a_k x = \underline{b}^{(k)}\right\}$ или $\overline{G}_k = \left\{x : a_k x = \overline{b}^{(k)}\right\}$.

Теорема 4. Если многогранник $M_k \neq \varnothing$ и активируемое ограничение $\overline{P}_k = \{x : a_k x \leqslant \overline{b}^{(k)}\}$ значимо для x^*_{k-1}, т. е. $J = \left\{j : a_k v_j(x^*_{k-1}) < 0\right\} \neq \varnothing$ и $a_k x^*_{k-1} > \overline{b}^{(k)}$, то в малой окрестности $U_\varepsilon(\lambda^*) = \{\lambda : 0 < \lambda^* - \lambda < \varepsilon\}$ вершина $\tilde{x}^*_k(\lambda)$, доставляющая целевой функции $L(x)$ максимум на $\tilde{M}_k(\lambda)$, может быть представлена в виде

$$\tilde{x}^*_k(\lambda) = x^*_{k-1} + \frac{\lambda - \lambda^*}{a_k v_{j_0}(x^*_{k-1})} v_{j_0}(x^*_{k-1}), \quad \lambda \in U_\varepsilon(\lambda^*), \tag{4}$$

где индекс j_0 определяется решающим правилом

$$\frac{cv_{j_0}(x^*_{k-1})}{a_k v_{j_0}(x^*_{k-1})} = \min_{j \in J} \frac{cv_j(x^*_{k-1})}{a_k v_j(x^*_{k-1})}. \tag{5}$$

Доказательство. В малой окрестности x^*_{k-1} при $\lambda \leqslant \lambda^* = a_k x^*_{k-1}$ вершинами $\tilde{M}_k(\lambda)$ являются точки пересечения рёбер $R_j(x^*_{k-1})$ многогранника M_{k-1} с гиперплоскостью $\tilde{G}_k(\lambda) = \{x : a_k x = \lambda\}$, и только они. Поэтому в согласии с теоремой 3 остаётся выбрать ту из них, в которой L достигает максимума. По построению $\tilde{x}_{k,j}(\lambda) = x^*_{k-1} + \dfrac{\lambda - \lambda^*}{a_k v_j(x^*_{k-1})} v_j(x^*_{k-1})$ является вершиной $\tilde{M}_k(\lambda)$, поскольку она лежит на ребре с направляющим вектором $v_j(x^*_{k-1})$ и в то же время оказывается на гиперплос-

кости $\tilde{G}_k(\lambda)$, что нетрудно проверить прямым расчётом:

$$a_k\tilde{x}_{k,j}(\lambda) = a_k\left(x^*_{k-1} + (\lambda-\lambda^*)\frac{v_j(x^*_{k-1})}{a_k v_j(x^*_{k-1})}\right) = a_k x^*_{k-1} + \lambda - \lambda^* = \lambda.$$

Целевая функция в этих вершинах принимает следующие значения

$$c\tilde{x}_{k,j}(\lambda) = cx^*_{k-1} + (\lambda-\lambda^*)\frac{cv_j(x^*_{k-1})}{a_k v_j(x^*_{k-1})}, \quad j \in J.$$

Здесь второе слагаемое неположительно, так как, согласно теореме 2, $cv_j(x^*_{k-1}) \leqslant 0$ и по условию $\lambda - \lambda^* \leqslant 0$, $a_k v_j(x^*_{k-1}) < 0$. Поэтому для всех $\lambda \in U_\varepsilon(\lambda^*)$ максимум L достигается на ребре с направляющим вектором $v_{j_0}(x^*_{k-1})$, для которого величина отношения $(cv_j(x^*_{k-1}))/(a_k v_j(x^*_{k-1}))$ минимальна, т. е. $\tilde{x}_{k,j_0}(\lambda) = \tilde{x}^*_k(\lambda)$. ■

Замечание 4. Полученный результат позволяет придать двойственным оценкам следующий смысл. Найдём двойственную оценку активируемого ограничения в вершине $\tilde{x}^*_k(\lambda)$ прямым расчётом с учётом её аналитического представления (4):

$$\frac{dL(\tilde{x}^*_k(\lambda))}{d\lambda} = \frac{d}{d\lambda}c\left(x^*_{k-1} + \frac{\lambda-\lambda^*}{a_k v_{j_0}(x^*_{k-1})}v_{j_0}(x^*_{k-1})\right) = \frac{cv_{j_0}(x^*_{k-1})}{a_k v_{j_0}(x^*_{k-1})}. \tag{6}$$

Таким образом, двойственная оценка активируемого ограничения по сути является решающим правилом (4), и наоборот, в чём и проявляется связь прямой и двойственной задач. Теперь формулы (4), (5) можно интерпретировать так: на текущей стадии активации значимого ограничения оптимальными будут точки ребра, вычисляющего двойственную оценку этого ограничения [13]. Правило (5) позволяет после нахождения оптимума вычислять двойственные переменные посредством последовательной псевдоактивации (назначения в качестве активируемого) каждого базового ограничения.

Далее для удобства принята сквозная нумерация ограничений: сначала координатные $(1,\ldots,n)$, а затем многофакторные $(n+1,\ldots,n+m)$.

Определение 3. Назовём базой вершины $\dot{x}$ совокупность номеров всех активных ограничений $B(\dot{x}) = \{i_k : a_{i_k}\dot{x} = b_{i_k},\ k = 1,\ldots,r \geqslant n\}$, т. е. тех, которым она удовлетворяет как равенствам (принадлежит гиперплоскостям), а сами ограничения (гиперплоскости) — базовыми. Величина $q = r - n + 1$ называется кратностью вершины. Однократная вершина — простая.

Активация ограничения в случае его значимости наряду с изменением оптимума влечёт соответствующую трансформацию базы. Локальный результат теоремы 4 допускает следующее расширение.

Теорема 5. База вершины $\tilde{x}^*_k(\lambda) = \tilde{x}_{k,j_0}(\lambda)$ сохраняется в пределах интервала $(\dot{\lambda}_{j_0}, \lambda^*)$, левая граница которого отвечает концевой точке $\dot{x}_{j_0}$ ребра $R_{j_0}(x^*_{k-1})$, т. е. $\dot{\lambda}_{j_0} = a_k\dot{x}_{j_0}$. Вершина $\dot{x}_{j_0}$, как смежная с x^*_{k-1}, вычисляется из расчёта максимального сдвига вдоль вектора $v_{j_0}(x^*_{k-1})$, допускаемого остальными ограничениями многогранника M_{k-1}, помимо базовых, а также противоположной гранью j_0-го базового условия.

Доказательство. Очевидно, что кандидатами на пересечение с ребром $R_{j_0}(x^*_{k-1})$ являются все активированные к настоящему моменту небазовые ограничения. В дополнение к этому необходимо учитывать также гарантированное пересечение данного ребра с противоположной гранью базового ограничения с номером j_0, поскольку из всех базовых гиперплоскостей направляющий вектор $v_{j_0}(x^*_{k-1})$ не ортогонален только

ей. Сообразно этим обстоятельствам параметры принятого для кандидата обозначения $\tilde{a}_s x = \tilde{b}^{(s)}$ конкретизируются по следующей схеме:

$$\left\{\begin{array}{l|l} \text{для координатных ограничений} & \text{для многофакторных ограничений} \\ v_{j_0}^{(s)}(x_{k-1}^*)>0 \Rightarrow \tilde{a}_s=e_s, \quad \tilde{b}^{(s)}=\overline{d}^{(s)}, & a_r v_{j_0}(x_{k-1}^*)>0 \Rightarrow \tilde{a}_s=a_r, \quad \tilde{b}^{(s)}=\overline{b}^{(r)}, \\ v_{j_0}^{(s)}(x_{k-1}^*)<0 \Rightarrow \tilde{a}_s=-e_s, \; \tilde{b}^{(s)}=-\underline{d}^{(s)}, & a_r v_{j_0}(x_{k-1}^*)<0 \Rightarrow \tilde{a}_s=-a_r, \; \tilde{b}^{(s)}=-\underline{b}^{(r)}, \\ s=1,\ldots,n, \;\; s \notin B(x_{k-1}^*) \setminus \{j_0\}\,, & r=1,\ldots,k, \; s=n+r \notin B(x_{k-1}^*) \setminus \{n+j_0\}\,. \end{array}\right.$$

Здесь e_s — единичный базисный вектор: $e_s^{(j)} = \delta_s^j$. Первая группа включает все небазовые координатные ограничения, а вторая — активированные к настоящему моменту небазовые многофакторные ограничения. Базовое ограничение с номером j_0 учитывается в той группе, к которой оно принадлежит.

Величина сдвига вдоль ребра $R_{j_0}(x_{k-1}^*)$ до пересечения с гиперплоскостью $G_s = \{x : \tilde{a}_s x = \tilde{b}^{(s)}\}$ в точке $x_{j_0,s}$ определяется соотношениями

$$x_{j_0,s} = x_{k-1}^* + t_{j_0,s} v_{j_0}(x_{k-1}^*), \;\; \tilde{a}_s \left(x_{k-1}^* + t_{j_0,s} v_{j_0}(x_{k-1}^*)\right) = \tilde{b}^{(s)}, \;\; t_{j_0,s} = \frac{\tilde{b}^{(s)} - \tilde{a}_s x_{k-1}^*}{\tilde{a}_s v_{j_0}(x_{k-1}^*)}.$$

Концевой точке ребра $\dot{x}_{j_0}$, сопряжённой с x_{k-1}^* по ребру $R_{j_0}(x_{k-1}^*)$, отвечает максимально возможный сдвиг в пределах многогранника M_{k-1}:

$$t_{j_0,s_0} = \min_{s \in S} t_{j_0,s}, \;\; S = \{1,\ldots,n+k-1\} \setminus B(x_{k-1}^*), \;\; \dot{x}_{j_0} = x_{k-1}^* + t_{j_0,s_0} v_{j_0}(x_{k-1}^*)$$

и $\dot{\lambda}_{j_0} = a_k \dot{x}_{j_0}$ — левый конец интервала стабильности базы $B(\tilde{x}_k^*(\lambda))$. ■

Таким образом, $\tilde{x}_k^*(\lambda)$ — оптимум L на $\tilde{M}_k(\lambda)$ для всех $\lambda \in [\dot{\lambda}_{j_0}, \lambda^*]$ и характер последующих действий определяется одной из двух ситуаций:

1) Если $\overline{b}^{(k)} \geqslant \dot{\lambda}_{j_0}$, то $x_k^* = \tilde{x}_k^*(\overline{b}^{(k)}) = x_{k-1}^* + \dfrac{\overline{b}^{(k)} - \lambda^*}{a_k v_{j_0}(x_{k-1}^*)} v_{j_0}(x_{k-1}^*)$. Преобразование $x_{k-1}^* \to x_k^*$ влечёт трансформацию базы $B(x_{k-1}^*) \to B(x_k^*) = \big(B(x_{k-1}^*) \setminus \{j_0\}\big) \cup \{k\}$, состоящую в замещении элемента j_0 на номер k, чем и завершается текущий шаг алгоритма.
2) В случае $\overline{b}^{(k)} < \dot{\lambda}_{j_0}$ активация ограничения продолжается относительно $\dot{x}_{j_0}$ с базой $B(\dot{x}_{j_0}) = \big(B(x_{k-1}^*) \setminus \{j_0\}\big) \cup \{s_0\}$ вплоть до занятия гиперплоскостью $\tilde{G}_k(\lambda)$ её геометрически окончательного положения.

На этом заканчивается построение вычислительных конструкций метода последовательной активации ограничений (МПАО), название которого происходит от характера производимых действий. В процессе активации ограничений теорема 1 выявляет несовместность системы условий задачи. Формулы теорем 4 и 5 вычисляют координаты оптимума и двойственные переменные. Неограниченность некоторых координат x^* означает неограниченность решения задачи по соответствующим переменным. В случае неединственности оптимума в интересах послеоптимизационного анализа теорема 2 позволяет перечислить смежные вершины в виде множества

$$\dot{X}^*(M) = \{x_m^*, \;\; \dot{x}_{m,j} = x_m^* + t_j v_j \in M : \; c v_j(x_m^*) = 0, \;\; j = 1,\ldots,n^*\}\,.$$

Каждая итерация заканчивается замещением в базе одного элемента другим, чему соответствует замена только одной строки матрицы нормалей активных ограничений.

Поэтому при нахождении направляющих векторов посредством инвертирования матриц использование факторизации существенно повышает быстродействие алгоритма.

Следуя первой ветви алгоритма с исключением переменных через равенства, по окончании решения задачи оптимальные значения исключенных переменных восстанавливаются по найденным на предварительном шаге их выражениям через действующие переменные.

Отличие ветви алгоритма без исключения равенств из системы условий состоит в том, что активация многофакторных ограничений производится в следующем порядке: сначала равенства как безусловно активные, затем неравенства. Порядок обращаемых матриц равен размерности пространства решений и, естественно, больше по сравнению с первой ветвью алгоритма. Зато объём других вычислений несколько уменьшается, поскольку по мере поступления равенств в базу промежуточной оптимальной вершины при активации ограничения в расчёт принимаются только те рёбра, которые удерживают последующие перемещения в гиперплоскостях равенств.

Подозрение на возможность зацикливания вычислений в МПАО возникает лишь относительно ситуации прохождения кратных вершин (определение 3), которая в результате накопления ошибок округления может как разрешиться сама собой, так и ложно возникнуть. В [3] в борьбе с зацикливанием предлагается делать ставку на регулярность метода. Следуя этой рекомендации, в случае неединственности оптимального ребра неоднозначность выбора в (5) разрешается по принципу наименьшего номера.

Процедура активации ограничения обладает регулярностью в том смысле, что в соответствии с теоремой 4 движение допускается вдоль направляющего вектора, образующего с нормальным вектором активируемого ограничения исключительно тупой угол. Это обстоятельство обеспечивает безаварийное прохождение вершин кратности больше 1, хотя МПАО настроен на простые вершины. Убедиться в этом помогает умозрительный пример 2-кратной вершины с применением к нему приёма «расклеивания» [14], суть которого состоит в представлении такой вершины в виде двух близко расположенных простых вершин за счёт малого изменения свободного члена со смещением «лишней» гиперплоскости вдоль её нормали во избежание сужения множества допустимых решений. В результате алгоритм если все-таки и пройдёт по фиктивному ребру, то всё равно выйдет на ненулевое оптимальное ребро. Однако тот же результат достижим при виртуальном расклеивании кратной вершины. Величина смещения вдоль фиктивного ребра равна 0, но, несмотря на сохранение координат промежуточной оптимальной вершины, произойдёт преобразование её базы, что означает фактический переход в структурно другую вершину. Повторение элементного состава базы означало бы возврат в ранее пройденную вершину, что возможно лишь при движении под запрещённым острым углом в обратном направлении. Это свидетельствует о безразличии МПАО к кратности вершин и невозможности зацикливания по этой причине. МПАО нечувствителен к вырожденности вершин в традиционной терминологии, легко распознаёт и разрешает случаи дублирования равенств и их маскировки совокупностью неравенств, что подтверждается экспериментальными расчётами.

В отличие от метода внутренней точки (МВТ), являющегося приближённым, МПАО — точный, подобно симплекс-методу (СМ). Однако теоретическая точность метода ещё не признак его безусловного превосходства, так как злую шутку может сыграть накопленная ошибка округления. В то же время приближённый метод может оказаться и точнее, и быстрее, что подтверждается практикой применения МВТ. Вычислительная эффективность МПАО определяется количеством шагов при активации

многофакторных ограничений. Её примитивная оценка сверху на базе подсчёта числа проходимых алгоритмом вершин промежуточных многогранников является ожидаемо экспоненциальной, и возможность её снижения вызывает законное сомнение.

2. Анализ результатов вычислительных экспериментов

МПАО реализован в MATLAB 7 R2011b программой `mpao`, которая решает задачу линейного программирования с двусторонними ограничениями, допуская неопределённость границ и отрицательность переменных. Условия-равенства оформляются равенством верхней и нижней границ допустимых изменений, в отдельную группу не выделяются и могут перемежаться неравенствами. Исходные данные передаются в программу массивом определённой структуры или текстовым файлом `trafaret` с вводом с клавиатуры только его номера при запуске программы. Данные из трафаретного файла считываются программой как строчные, что обеспечивает отсутствие жёстких требований к формату их расположения в поле ввода и визуализацию исходных данных сообразно вкусам пользователя. В случаях несовместности, неограниченности, неединственности решения выдаются соответствующие сообщения. Всплывающие меню с активными кнопками обеспечивают управление функциональными возможностями программы в части её запуска, визуализации исходных данных и полученных результатов, вычисления альтернативных решений в случае неединственности.

Рассмотрим результаты применения программ `mpao` и `linprog` библиотеки MATLAB в опциях крупномасштабного алгоритма LSA (МВТ) и среднемасштабного алгоритма MSA (СМ) на простом примере

$$L = -x^{(1)} + x^{(2)} \to \min,$$
$$\begin{cases} -x^{(1)} + x^{(2)} \geqslant 1, \\ -2x^{(1)} + x^{(2)} \leqslant 2, \\ 3x^{(1)} + x^{(2)} \leqslant 3 \end{cases}$$

с допустимыми решениями в треугольнике с вершинами A(0,5, 1,5), B(0,2, 2,4), C(-1, 0) и оптимумом на ребре AC.

В программе `linprog` не предусмотрено выявление неединственности решения, хотя уже сам факт возникновения такой ситуации в практических задачах может быть весьма весомым. Опция LSA вышла на внутреннюю точку $x^* = (-0{,}2743, 0{,}7257)$ ребра AC, а MSA — на вершину $x^* = (-1, 0)$ и там остановилась.

Программа `mpao` нашла обе оптимальные вершины $x_1^* = (0{,}5, 1{,}5)$ и $x_2^* = (-1, 0)$, т. е. всё ребро AC. Вычисленные двойственные переменные в обоих решениях имеют одинаковые соответствующие значения $(1, 0, 0)$, что объясняется следующим обстоятельством. В первом решении активными являются первое и третье ограничения, а во втором — первое и второе. Превратим активные ограничения в свободные и объединим их правые части $\lambda^{(k)}$ в вектор $\lambda = (\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)})$. При условии сохранения базы оптимальной вершины (теорема 5) прямыми выкладками получаем $L(\tilde{x}_1^*(\lambda)) = L(\tilde{x}_2^*(\lambda)) = \lambda^{(1)}$. Обозначив оптимум двойственных переменных $y^* = (y^{(1)^*}, y^{(2)^*}, y^{(3)^*})$, следуя формуле (6) и [14], находим

$$y^{(1)^*} = \frac{\partial L(\tilde{x}_1^*(\lambda))}{\partial \lambda^{(1)}} = \frac{\partial L(\tilde{x}_2^*(\lambda))}{\partial \lambda^{(1)}} = 1,\ y^{(k)^*} = \frac{\partial L(\tilde{x}_1^*(\lambda))}{\partial \lambda^{(k)}} = \frac{\partial L(\tilde{x}_2^*(\lambda))}{\partial \lambda^{(k)}} = 0,\ k = 2, 3,$$

т. е. у двойственной задачи оптимум $y^* = (1, 0, 0)$ единственен, что и подтверждается её непосредственным решением.

Замена первого ограничения на $-x^{(1)}+x^{(2)} \geqslant 3$ превращает систему условий в несовместную. Реакция `linprog`: решение не найдено. Диагностика `mpao`: совместность нарушает ограничение 3 (правостороннее). Другой порядок активации ограничений может изменить номер нарушителя.

Изъятие первого неравенства влечёт неограниченность решения. Сообщение `linprog`: решение не ограничено, $x^* = (2{,}46 \cdot 10^{31}, -7{,}39 \cdot 10^{31})$, $L = -9{,}85 \cdot 10^{31}$. Ответ `mpao`: решение неограничено по переменным $x^{(1)}$ и $x^{(2)}$, $L = -\infty$.

Масштабная апробация программы `mpao` производилась на базе задач, созданных с помощью специальной модификации `mpao`, для которой исходными данными являются количество переменных и ограничений, а также наличие или отсутствие требования совместности системы условий. Используя генератор случайных чисел, эта программа выдаёт коэффициенты целевой функции и ограничений, а также лимиты ограничений и переменных. В таблице приведено время счёта для задач с размерностями в диапазоне $10 \times 10 \div 1000 \times 1000$ для программ `mpao` и `linprog` на одноядерном ПК с частотой процессора 1,73 ГГц и ОЗУ 1 Гб. Столь маломощный компьютер был выбран умышленно для обеспечения ощутимости времени счёта на задачах малой размерности.

Размер	`mpao`			LSA			MSA		
$m \times n$	мин.	сред.	макс.	мин.	сред.	макс.	мин.	сред.	макс.
10×10	0,00	0,00	0,08	0,03	0,04	0,67	0,01	0,025	0,562
20×20	0,00	0,01	0,05	0,03	0,05	0,09	0,06	0,074	0,219
30×30	0,03	0,04	0,12	0,11	0,13	0,78	0,16	0,185	0,719
40×40	0,09	0,11	0,12	0,12	0,13	0,17	0,34	0,369	0,454
50×50	0,13	0,14	0,2	0,12	0,14	0,20	0,73	0,758	0,875
80×80	0,73	0,76	0,89	0,40	0,44	0,56	2,89	3,067	3,485
100×100	1,44	1,47	1,52	0,76	0,79	0,86	7,03	7,333	7,593
200×200	16,4	16,6	16,9	6,09	6,21	6,31	116	116,8	118,3
300×300	60,2	60,6	60,9	6,75	6,98	7,61	310	310,5	311,1
300×500	327	328	329	21,8	21,9	22,4	617	618,3	622,4
500×300	98,9	101	101	17,1	17,3	17,8	874	874,7	875,3
500×500	437	442	446	36,4	36,7	37,0	2346	2351	2356
500×1000	3393	3410	3433	447	471	516	5902	5904	5906
1000×500	805	851	942	45,5	46,5	47,7	10895	11113	11505
1000×1000	5854	5862	5869	328	333	345	46091	48961	51223

П р и м е ч а н и е. Указаны минимальное (мин.), максимальное (макс.) и усреднённое (сред.) время счёта в секундах по 1000 прогонов для каждой из малоразмерных задач с постепенным снижением числа прогонов до трёх по мере роста размерности. Усреднённое время рассчитано как среднеарифметическое по числу прогонов.

В экспериментальных расчётах последовательность активации ограничений заметно влияла на количество итераций, и это обстоятельство логично использовать с максимальной выгодой. Пока что для повышения эффективности алгоритма порядок активации устанавливается исходя из просто реализуемых и по возможности малообременительных правил. Одним из них является выбор кандидата на активацию с максимальной значимостью. Под мерой значимости может пониматься, например, возможность наибольшего снижения максимума целевой функции при активации ограничения, что подразумевает выявление кандидата на активацию на каждом шаге алгоритма. Активацию ограничений можно производить также в порядке убывания угла между нормалями целевой функции и ограничений. Детали такой стратегии устанавливаются один раз изначально, основанием для этого служит следующее соображение. В E^2 одна из нормалей базовой гиперплоскости оптимальной вершины составляет с

нормалью целевой функции минимальный угол. Хотя уже в E^3 это вовсе не обязательно, результаты экспериментальных расчётов показывают, что такая последовательность активации, не будучи наилучшей, сокращает количество итераций на десятки процентов и потому вполне оправдана. В случае нахождения оптимальной стратегии активации ограничений традиционным неопровержимым аргументом против принципиальной возможности понижения экспоненциальной оценки числа итераций МПАО будет контрпример.

Анализ полученных результатов свидетельствует, что, в противоположность MSA, быстродействие `mpao` и LSA в большей мере зависит от числа переменных, нежели от количества ограничений. Это обстоятельство обеспечивает первой ветви алгоритма (с исключением части переменных через равенства) вычислительное преимущество по сравнению со второй ветвью (решение задачи непосредственно с равенствами). На малых размерностях до 50×50 уверенно лидирует `mpao`, а далее при стабильном удержании ею второго места быстро растущее превосходство переходит к опции LSA, которая демонстрирует слабую зависимость времени счёта от размерности задач. На размерности 1000×1000 MSA многократно уступает LSA и `mpao`. На ПК с двухъядерным процессором с частотой 2 ГГц и ОЗУ 2 Гб вычисления ускорились примерно на 10 % с сохранением соотношения величин этого показателя для всех трёх программ.

В условиях наличия высокопроизводительных компьютеров быстродействие далеко не всегда является основным критерием эффективности алгоритма и соответствующего программного продукта, на первое место могут выдвигаться их функциональные и сервисные возможности.

Заключение

Сопоставление СМ и МПАО приводит к следующим выводам. Оба метода — точные и конечные. Решая задачу на максимум, СМ на каждом шаге пытается увеличить текущее значение целевой функции, продвигаясь вдоль одной из координатных осей в границах МДР, т. е. максимум достигается снизу. Для превращения неравенств в равенства и нахождения начальной точки вводятся искусственные переменные, что формально повышает размерность задачи.

МПАО, напротив, предпочитает сокращать количество переменных путём их исключения через посредство равенств, поскольку его вычислительная эффективность обратно пропорциональна числу неизвестных. Последовательно активируя ограничения, алгоритм усекает начальный гиперпараллелепипед до заданного МДР, сужая область допустимых решений и сохраняя состояние оптимальности. В результате этого максимальное значение целевой функции на каждом шаге уменьшается или, по крайней мере, не увеличивается и приближение к максимуму выглядит как спуск сверху. Начальная точка маршрута определяется тривиально, и последующее движение к оптимальной вершине происходит по траектории, представляющей собой ломаную прямую, вершинами которой являются оптимальные вершины промежуточных многогранников. Эта траектория пролегает вне МДР; если в ходе активации ограничений не выявлена их несовместность, то его первой же захваченной алгоритмом точкой как раз и будет искомая оптимальная вершина. В связи с этим напрашиваются такие параллели: симплекс метод — метод внутренней точки, метод последовательной активации ограничений — метод внешней точки. В отличие от геометрического метода, одолевающего задачи максимум с тремя переменными с визуализацией МДР, МПАО, базируясь на геометрических представлениях, решает задачи любой размерности. МПАО активно использует органичную связь прямой и двойственной задач. Это проявляется в том,

что решающее правило выбора оптимального ребра (5) вычисляет двойственную оценку ограничения на текущей стадии его активации. После нахождения оптимума задачи посредством псевдоактивации каждого из базовых ограничений решающее правило находит оптимальное решение двойственной задачи. Последнее действие вычислительно малообременительно, поскольку к его началу известны направляющие векторы рёбер многогранника допустимых решений, выходящих из оптимальной вершины, и активные в ней ограничения. Таким образом, в МПАО двойственность играет решающую роль на всех стадиях вычислительного процесса.

Анализ результатов экспериментальных расчётов свидетельствует о теоретической и практической состоятельности МПАО и превосходстве в быстродействии программы `mpao` над опцией симплекс-метода программы `linprog`, растущим с размерностью задач. Потенциал программы позволяет: констатировать неограниченность целевой функции и указывать ответственные за это переменные; выявлять ограничения, вызывающие несовместность системы условий; в отличие от `linprog`, в случае неединственности описывать структуру множества решений в виде смежных оптимальных вершин; определять чувствительность оптимума относительно активных ограничений посредством двойственных оценок; подобно прожектору, высвечивать структуру МДР в окрестности оптимума, что в случае неполной достоверности исходных данных имеет практическое значение. Резервы повышения быстродействия МПАО ещё не исчерпаны, но уже сейчас всего лишь с экспериментальной программой `mpao` он уверенно занимает среднее положение между СМ и МВТ и по ряду параметров более информативен. Стремительное продвижение к оптимуму МВТ вне конкуренции и точным методом пока недостижимо.

ЛИТЕРАТУРА

1. *Канторович Л. В.* Математические методы в организации и планировании производства. Л.: Изв. Ленингр. гос. ун-та, 1939. 67 с.
2. *Данциг Дж.* Линейное программирование, его применение и обобщения. М.: Прогресс, 1966. 600 с.
3. *Юдин Д. Б., Гольштейн Е. Г.* Линейное программирование: теория, методы и приложения. М.: Наука, 1969. 424 с.
4. *Дикин И. И.* Итеративное решение задач линейного и квадратичного программирования // Докл. АН СССР. 1967. Т. 174. С. 747–748.
5. *Хачиян Л. Г.* Полиномиальный алгоритм в линейном программировании // Докл. АН СССР. 1979. Т. 244. С. 1093–1096.
6. *Зоркальцев В. И.* Двойственные алгоритмы внутренних точек // Изв. вузов. Математика. 2011. № 4. С. 33–53.
7. *Зоркальцев В. И., Медвежонков Д. С.* Численные эксперименты с вариантами алгоритмов внутренних точек на нелинейных задачах потокораспределения // Управление большими системами. Ин-т систем энергетики им. Л. А. Мелентьева СО РАН, 2013. Вып. 46. С. 68–87.
8. *Медвежонков Д. С.* Экспериментальные исследования алгоритмов внутренних точек на нелинейных задачах потокораспределения // Вестник Бурятского гос. ун-та. 2013. № 9. С. 12–16.
9. *Вылегжанин О. Н., Шкатова Г. И.* Решение задачи линейного программирования с использованием оператора-проектора // Изв. Томского политехн. ун-та. 2009. Т. 314. № 5. С. 37–40.

10. *Бахшиян Б. Ц., Гориянов А. В.* Скелетный алгоритм решения задачи линейного программирования и его применение для решения задач оценивания // Вестник МАИ. 2008. Т. 15. № 2. С. 5–16.
11. *Гордуновский В. М.* Метод экспоненциальной аппроксимации для линейного программирования. М.: Эдитус, 2013. 32 с.
12. *Колосов В. С.* Конечный параметрический метод решения задачи линейного программирования // Науч. труды МЛТИ. 1978. Вып. 103. С. 197–198.
13. *Колосов В. С.* Роль двойственных оценок в параметрическом методе решения задачи линейного программирования // Науч. труды МЛТИ. 1986. Вып. 183. С. 51–54.
14. *Карманов В. Г.* Математическое программирование. М.: ФИЗМАТЛИТ, 2004. 264 с.

REFERENCES

1. *Kantorovich L. V.* Matematicheskie metody v organizacii i planirovanii proizvodstva [Mathematical Methods in the Organization and Planning of Production]. Leningrad, News of Leningrad State University, 1939. 67 p. (in Russian)
2. *Dantzig J.* Linejnoe programmirovanie, ego primenenie i obobshcheniya [Linear Programming, its Application and Generalizations]. Moscow, Progress Publ., 1966. 600 p. (in Russian)
3. *Yudin D. B. and Golstein E. G.* Linejnoe programmirovanie: teoriya, metody i prilozheniya [Linear Programming: Theory, Methods and Applications]. Moscow, Nauka Publ., 1969. 424 p. (in Russian)
4. *Dikin I. I.* Iterativnoe reshenie zadach lineynogo i kvadratichnogo programmirovaniya [Iterative solution of problems of linear and quadratic programming]. Report AS USSR, 1967, vol. 174, pp. 747–748. (in Russian)
5. *Khachiyan L. G.* Polinomial'nyy algoritm v lineynom programmirovanii [Polynomial algorithm in linear programming]. Report AS USSR, 1979, vol. 244, pp. 1093–1096. (in Russian)
6. *Zorkaltsev V. I.* Dvoystvennye algoritmy vnutrennikh tochek [Dual algorithms of interior-points]. Iz. VUZ, Mathematics, 2011, no. 4 pp. 33–53. (in Russian)
7. *Zorkaltsev V. I. and Medvejonkov D. S.* Chislennye eksperimenty s variantami algoritmov vnutrennikh tochek na nelineynykh zadachakh potokoraspredeleniya [Computational experiments with variants of interior-point algorithms for nonlinear flow distribution problems]. Institute of Energy Systems of SB RAS, 2013, no. 46, pp. 68–87. (in Russian)
8. *Medvejonkov D. S.* Eksperimental'nyye issledovaniya algoritmov vnutrennikh tochek na nelineynykh zadachakh potokoraspredeleniya [Experimental researches of interior- point algorithms for solution of flow distribution nonlinear problems]. Bulletin of the Buryat State University, 2013, no. 9, pp. 12–16. (in Russian)
9. *Vylegzhanin O N. and Skatova G. I.* Reshenie zadachi lineynogo programmirovaniya s ispol'zovaniem operatora-proektora [The solution of the linear programming problem using the operator- projector]. News Tomsk Polytechnic University, 2009, vol. 314, no. 5, pp. 37–40. (in Russian)
10. *Bakhshiyan B. Ts. and Gurianov A. V.* Skeletnyy algoritm resheniya zadachi lineynogo programmirovaniya i ego primenenie dlya resheniya zadach otsenivaniya [The skeletal algorithm for solving the linear programming problem and its application for solving estimation problems]. Bulletin MAI, 2008, vol. 15, no. 2, pp. 5–16. (in Russian)
11. *Gordunovski V. M.* Metod ehksponencial'noj approksimacii dlya linejnogo programmirovaniya [The Method of Exponential Approximation for Linear Programming]. Moscow, Editus Publ., 2013. 32 p. (in Russian)

12. *Kolosov V. S.* Konechnyy parametricheskiy metod resheniya zadachi lineynogo programmirovaniya [The finite parametric method for solving linear programming problem]. Scientific Works MLTI, 1978, vol. 103, pp. 197–198. (in Russian)
13. *Kolosov V. S.* Rol' dvoystvennykh otsenok v parametricheskom metode resheniya zadachi lineynogo programmirovaniya [The role of dual estimates in the parametric method for solving the linear programming problem]. Scientific Works MLTI, 1986, vol. 183, pp. 51–54. (in Russian)
14. *Karmanov V. G.* Matematicheskoe programmirovanie [Mathematical Programming]. Moscow, FIZMATLIT Publ., 2004. 264 p.

СВЕДЕНИЯ ОБ АВТОРАХ

АНОХИН Михаил Игоревич — кандидат физико-математических наук, старший научный сотрудник Института проблем информационной безопасности Московского государственного университета имени М. В. Ломоносова, г. Москва.
E-mail: **anokhin@mccme.ru**

АРТЕМОВА Наталья Александровна — ассистент Саратовского национального исследовательского государственного университета имени Н. Г. Чернышевского, г. Саратов. E-mail: **NatalyaKorArt@ya.ru**

БЕЛИМ Сергей Викторович — доктор физико-математических наук, профессор, заведующий кафедрой информационной безопасности Омского государственного университета им. Ф. М. Достоевского, г. Омск. E-mail: **belimsv@omsu.ru**

БОГАЧЕНКО Надежда Федоровна — кандидат физико-математических наук, доцент, доцент кафедры информационной безопасности Омского государственного университета им. Ф. М. Достоевского, г. Омск. E-mail: **nfbogachenko@mail.ru**

ГЕЙДАРОВ Полад Шахмалы оглы — кандидат технических наук, доцент, ведущий научный сотрудник Института систем управления НАН Азербайджана, г. Баку. E-mail: **plbaku2010@gmail.com**

ДЕНИСОВ Олег Викторович — кандидат физико-математических наук, доцент, ООО «Инновационные телекоммуникационные технологии», г. Москва.
E-mail: **denisovOleg@yandex.ru**

ИДРИСОВА Валерия Александровна — аспирантка Института математики им. С. Л. Соболева СО РАН, лаборантка лаборатории алгоритмики Новосибирского государственного университета, г. Новосибирск. E-mail: **vitkup@math.nsc.ru**

КОЛОСОВ Вадим Сергеевич — кандидат технических наук, старший научный сотрудник, г. Москва. E-mail: **vs.kolosov@yandex.ru**

ЛЕБЕДЕВ Филипп Владимирович — специалист по информационной безопасности компании ООО «АСП Лабс», г. Москва. E-mail: **plebedev@asplabs.ru**

МОНАХОВА Эмилия Анатольевна — кандидат технических наук, доцент, старший научный сотрудник Института вычислительной математики и математической геофизики СО РАН, г. Новосибирск. E-mail: **emilia@rav.sscc.ru**

ОБЗОР Анастасия Александровна — студентка кафедры компьютерной математики и программирования Омского государственного университета им. Ф. М. Достоевского, г. Омск. E-mail: **obzor2503@gmail.com**

РОМАНЬКОВ Виталий Анатольевич — доктор физико-математических наук, профессор, заведующий кафедрой компьютерной математики и программирования Омского государственного университета им. Ф. М. Достоевского, г. Омск.
E-mail: **romankov48@mail.ru**

РЯЗАНОВ Юрий Дмитриевич — доцент, доцент Белгородского государственного технологического университета им. В. Г. Шухова, г. Белгород.
E-mail: **razanov.yd@bstu.ru**