МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ПРИКЛАДНАЯ ДИСКРЕТНАЯ МАТЕМАТИКА

Научный журнал

2017 № 37

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 20.09.2017. Формат $60 \times 84\frac{1}{8}$. Усл. п. л. 14,55. Тираж 300 экз.
Заказ № 2743. Цена свободная. Дата выхода в свет 04.10.2017.

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

СОДЕРЖАНИЕ

CONTENTS

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

УДК 512.541, 512.552, 512.711

О ДВУХ ОПРЕДЕЛЕНИЯХ СТЕПЕНИ ФУНКЦИИ НАД АССОЦИАТИВНЫМ КОММУТАТИВНЫМ КОЛЬЦОМ[1]

М. И. Анохин

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

Пусть R — ассоциативное коммутативное кольцо и $\varphi\colon R^m \to R$, где $m \geqslant 0$. Обозначим через $\deg_{\Pi}\varphi$ наименьшее число $n \geqslant -1$, такое, что φ представима многочленом степени n от m переменных над R. (Степенью нулевого многочлена считаем -1.) Пусть также $\deg_{\mathrm{RM}}\varphi$ обозначает наименьшее число $n \geqslant -1$, такое, что $\partial_{v_1}\dots\partial_{v_{n+1}}\varphi = 0$ для всех $v_1,\dots,v_{n+1} \in R^m$. Здесь $(\partial_v\psi)(x) = \psi(x+v) - \psi(x)$ для любых $v, x \in R^m$ и любой функции $\psi\colon R^m \to R$. Если такого числа n не существует, то полагаем соответственно $\deg_{\Pi}\varphi = \infty$ или $\deg_{\mathrm{RM}}\varphi = \infty$. В работе рассматривается проблема характеризации класса $\mathfrak{D}$ всех ассоциативных коммутативных колец R, таких, что эти степени совпадают для функций над R, т. е. $\deg_{\Pi}\varphi = \deg_{\mathrm{RM}}\varphi$ для всех $m \geqslant 0$ и всех функций $\varphi\colon R^m \to R$. Проблема решается в случае, когда аддитивная группа $\mathcal{R}$ кольца R принадлежит некоторым широким классам абелевых групп. Основные результаты: 1) если $\mathcal{R}$ периодична или конечно порождена, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong \mathbb{Z}/d\mathbb{Z}$ для некоторого свободного от квадратов числа $d \geqslant 1$; 2) если $\mathcal{R}$ не редуцирована, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong (\mathbb{Z}/d\mathbb{Z}) \oplus \mathbb{Q}$ для некоторого свободного от квадратов числа $d \geqslant 1$; 3) если $\mathcal{R}$ является прямой суммой подгрупп ранга 1, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong \mathbb{Z}/d\mathbb{Z}$ или $R \cong (\mathbb{Z}/d\mathbb{Z}) \oplus \mathbb{Q}$ для некоторого свободного от квадратов числа $d \geqslant 1$; 4) если $\mathcal{R}$ редуцирована и копериодична, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong \prod\limits_{p\in P}(\mathbb{Z}/p\mathbb{Z})$ для некоторого множества P простых чисел. Доказательство этих результатов основано на том факте, что любое кольцо из $\mathfrak{D}$ является E-кольцом.

Ключевые слова: *ассоциативное кольцо, коммутативное кольцо, абелева группа, аддитивная группа кольца, многочлен, степень функции, E-кольцо, формула Ньютона.*

DOI 10.17223/20710410/37/1

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

ON THE TWO DEFINITIONS OF DEGREE OF A FUNCTION OVER AN ASSOCIATIVE, COMMUTATIVE RING

M. I. Anokhin

Information Security Institute of Lomonosov University, Moscow, Russia

E-mail: anokhin@mccme.ru

Let R be an associative, commutative ring and let $\varphi\colon R^m \to R$, where $m \geqslant 0$. Denote by $\deg_\Pi \varphi$ the smallest integer $n \geqslant -1$ such that φ can be represented by an m-variate polynomial of degree n over R. (By convention, the degree of the zero polynomial is -1.) Also, let $\deg_{\mathrm{RM}} \varphi$ denote the smallest integer $n \geqslant -1$ such that $\partial_{v_1} \dots \partial_{v_{n+1}} \varphi = 0$ for all $v_1, \dots, v_{n+1} \in R^m$. Here $(\partial_v \psi)(x) = \psi(x+v) - \psi(x)$ for any $v, x \in R^m$ and any function $\psi\colon R^m \to R$. If no such integer n exists, then we put $\deg_\Pi \varphi = \infty$ or $\deg_{\mathrm{RM}} \varphi = \infty$, respectively. In this paper, we study the problem of characterizing the class $\mathfrak{D}$ of all associative, commutative rings R such that these degrees coincide for functions over R, i.e., $\deg_\Pi \varphi = \deg_{\mathrm{RM}} \varphi$ for all $m \geqslant 0$ and all functions $\varphi\colon R^m \to R$. We solve this problem when the additive group $\mathcal{R}$ of the ring R belongs to some large classes of abelian groups. Namely, our main results are as follows: 1) if $\mathcal{R}$ is torsion or finitely generated, then $R \in \mathfrak{D}$ if and only if $R \cong \mathbb{Z}/d\mathbb{Z}$ for some square-free integer $d \geqslant 1$; 2) if $\mathcal{R}$ is not reduced, then $R \in \mathfrak{D}$ if and only if $R \cong (\mathbb{Z}/d\mathbb{Z}) \oplus \mathbb{Q}$ for some square-free integer $d \geqslant 1$; 3) if $\mathcal{R}$ is a direct sum of rank 1 subgroups, then $R \in \mathfrak{D}$ if and only if $R \cong \mathbb{Z}/d\mathbb{Z}$ or $R \cong (\mathbb{Z}/d\mathbb{Z}) \oplus \mathbb{Q}$ for some square-free integer $d \geqslant 1$; 4) if $\mathcal{R}$ is reduced and cotorsion, then $R \in \mathfrak{D}$ if and only if $R \cong \prod\limits_{p \in P} (\mathbb{Z}/p\mathbb{Z})$ for some set P of prime numbers. The proof of these results is based on the fact that any ring in $\mathfrak{D}$ is an E-ring.

Keywords: *associative ring, commutative ring, Abelian group, additive group of a ring, polynomial, degree of a function, E-ring, Newton's formula.*

Введение

В настоящей работе аддитивная группа произвольного кольца обозначается рукописным вариантом буквы, обозначающей это кольцо. Например, $\mathcal{Q}$, $\mathcal{Z}$ и $\mathcal{Z}_n$ (где n — целое положительное число) — это аддитивные группы поля $\mathbb{Q}$ рациональных чисел, кольца $\mathbb{Z}$ целых чисел и кольца $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ соответственно. Для произвольного $k \in \mathbb{Z}$ положим $\mathbb{N}_k = \{n \in \mathbb{Z} : n \geqslant k\}$. Пусть также $\mathbb{P}$ — множество всех простых чисел. Все абелевы группы записываются аддитивно. Будем придерживаться соглашения о том, что сумма пустого множества элементов абелевой группы (в частности, аддитивной группы кольца) равна 0. Аналогично, произведение пустого набора элементов ассоциативного кольца с единицей считается равным 1; в частности, $r^0 = 1$ для любого элемента r такого кольца.

Для $m \in \mathbb{N}_0$ через X^m обозначается прямое произведение m экземпляров множества, группы или кольца X. В частности, X^0 состоит из одного элемента. Пусть $(X_i : i \in I)$ — семейство колец или абелевых групп. Тогда через $\prod\limits_{i \in I} X_i$ и $\bigoplus\limits_{i \in I} X_i$ будем обозначать соответственно прямое произведение (называемое также полной прямой суммой) и внешнюю прямую сумму этого семейства. Разумеется, если I конечно, то $\prod\limits_{i \in I} X_i = \bigoplus\limits_{i \in I} X_i$. Символом $\oplus$ обозначается бинарная операция прямой суммы (как внешней, так и внутренней). Операция $\prod$ используется и для произвольных (не обязательно

абелевых) групп, записываемых мультипликативно; для них она называется операцией декартова произведения. Элементы $\prod\limits_{i\in I} X_i$ обозначаются $(x_i : i \in I)$, где $x_i \in X_i$ для всех $i \in I$.

Пусть $m \in \mathbb{N}_0$ и R — ассоциативное коммутативное кольцо (вообще говоря, без единицы). Для каждого $n \in \mathbb{N}_{-1}$ обозначим через $\Pi_n(R, m)$ множество всех функций из R^m в R, представимых многочленами степени не более n от m переменных над кольцом R. Нулевому многочлену приписываем степень -1. Другими словами, $\Pi_n(R, m)$ состоит из всевозможных сумм функций вида $(x_1, \ldots, x_m) \mapsto r x_{i_1} \ldots x_{i_s}$ $(x_1, \ldots, x_m \in R)$, где $r \in R$, $0 \leqslant s \leqslant n$ и $1 \leqslant i_1 \leqslant \ldots \leqslant i_s \leqslant m$. В частности, $\Pi_{-1}(R, m) = \{0\}$ (функция, тождественно равная 0, также обозначается через 0). Из определения следует, что $\Pi_n(R, m) \subseteq \Pi_{n+1}(R, m)$ для всех $n \in \mathbb{N}_{-1}$. Пусть также $\Pi(R, m) = \bigcup\limits_{n=-1}^{\infty} \Pi_n(R, m)$. Естественно определить степень произвольной функции $\varphi\colon R^m \to R$ как $\min\{n \in \mathbb{N}_{-1} : \varphi \in \Pi_n(R, m)\}$, если $\varphi \in \Pi(R, m)$, и ∞ в противном случае. Эту степень обозначим через $\deg_\Pi \varphi$.

С другой стороны, для каждого $n \in \mathbb{N}_{-1}$ можно определить множество $\mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$, следуя [1]. Подчеркнём, что в определении этого множества участвует не кольцо R, а лишь его аддитивная группа $\mathcal{R}$. Приведём определение множеств $\mathrm{RM}_n(G, A)$ для произвольной группы G и произвольной абелевой группы A (см. также [2]; там эти множества обозначаются через $\Phi_n(G, A)$). Напомним, что *производной* функции $\varphi\colon G \to A$ по направлению $g \in G$ называется функция $\partial_g\varphi\colon G \to A$, определённая равенством $(\partial_g\varphi)(x) = \varphi(xg) - \varphi(x)$ для каждого $x \in G$. Множества $\mathrm{RM}_n(G, A)$ определяются индукцией по n следующим образом:

$$\mathrm{RM}_{-1}(G, A) = \{0\},$$
$$\mathrm{RM}_n(G, A) = \{\varphi\colon G \to A : \forall\, g \in G\ \ (\partial_g\varphi \in \mathrm{RM}_{n-1}(G, A))\} \quad \text{при } n \in \mathbb{N}_0.$$

Отметим, что в работе [1] таким образом определены множества $\mathrm{RM}_n(G, A)$ $(n \in \mathbb{N}_{-1})$ в случае, когда G и A — конечные абелевы группы. Легко видеть, что

$$\mathrm{RM}_n(G, A) = \{\varphi\colon G \to A : \forall\, g_1, \ldots, g_{n+1} \in G\ (\partial_{g_1} \ldots \partial_{g_{n+1}}\varphi = 0)\},$$

поэтому $\mathrm{RM}_n(G, A) \subseteq \mathrm{RM}_{n+1}(G, A)$ для каждого $n \in \mathbb{N}_{-1}$. Положим также $\mathrm{RM}(G, A) = \bigcup\limits_{n=-1}^{\infty} \mathrm{RM}_n(G, A)$.

Обозначение $\mathrm{RM}_n(G, A)$ указывает на связь этого множества с кодом Рида — Маллера порядка n. Действительно, p-ичный код Рида — Маллера порядка n и длины p^m (где p — простое число) может быть определён как $\Pi_n(\mathbb{Z}_p, m)$. В то же время известно, что $\Pi_n(\mathbb{Z}_p, m) = \mathrm{RM}_n(\mathcal{Z}_p^m, \mathcal{Z}_p)$ для всех $p \in \mathbb{P}$, $m \in \mathbb{N}_0$ и $n \in \mathbb{N}_{-1}$ ([3, свойство B7], а также следствие 1 ниже). Можно также рассматривать $\mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$ как множество функций степени не более n, альтернативное $\Pi_n(R, m)$. Соответствующую степень произвольной функции $\varphi\colon R^m \to R$ обозначим через $\deg_{\mathrm{RM}} \varphi$, а именно:

$$\deg_{\mathrm{RM}} \varphi = \begin{cases} \min\{n \in \mathbb{N}_{-1} : \varphi \in \mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})\}, & \text{если } \varphi \in \mathrm{RM}(\mathcal{R}^m, \mathcal{R}), \\ \infty & \text{в противном случае.} \end{cases}$$

Замечание 1. Непосредственно проверяется, что если $\pi \in \Pi_n(R, m)$, где $n \in \mathbb{N}_0$, то $\partial_v\pi \in \Pi_{n-1}(R, m)$ при любом $v \in \mathcal{R}^m$. Поэтому индукция по n показывает, что $\Pi_n(R, m) \subseteq \mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$ для всех $n \in \mathbb{N}_{-1}$. Следовательно, $\deg_{\mathrm{RM}} \varphi \leqslant \deg_\Pi \varphi$ для любой функции $\varphi\colon R^m \to R$.

В настоящей работе изучается следующий вопрос: для каких ассоциативных коммутативных колец R равенство $\deg_\Pi \varphi = \deg_{\mathrm{RM}} \varphi$ верно для всех $m \in \mathbb{N}_0$ и всех функций $\varphi\colon R^m \to R$? Эквивалентным образом этот вопрос можно сформулировать так: для каких ассоциативных коммутативных колец R равенство $\Pi_n(R, m) = \mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$ справедливо при любых $m \in \mathbb{N}_0$ и $n \in \mathbb{N}_{-1}$? Класс всех таких колец обозначим через $\mathfrak{D}$. Ранее было известно, что $\mathbb{Z}_p \in \mathfrak{D}$ для всех простых чисел p ([3, свойство B7]; для $p = 2$ этот факт отмечается в разд. 2 работы [1]). Отметим также результат А. В. Черемушкина [4, теорема 6]: для любой функции $\pi \in \Pi(\mathbb{Z}_{p^k}, m)$, где $p \in \mathbb{P}$ и $k, m \in \mathbb{N}_1$, имеет место равенство $\deg_\Pi \pi = \deg_{\mathrm{RM}} \pi$. Это утверждение эквивалентно тому, что $\Pi_n(\mathbb{Z}_{p^k}, m) = \mathrm{RM}_n(\mathcal{Z}^m_{p^k}, \mathcal{Z}_{p^k}) \cap \Pi(\mathbb{Z}_{p^k}, m)$ для всех $n \in \mathbb{N}_{-1}$ и всех p, k и m указанного вида. Отметим, что из этого результата, следствия 1, замечания 5 и эквивалентности из доказательства леммы 3 вытекает, что данный результат верен для колец $\mathbb{Z}_d$ и $\mathbb{Z}_d \oplus \mathbb{Q}$ при всех $d \in \mathbb{N}_1$, а также для колец вида $\prod\limits_{p \in P} \mathbb{Z}_{p^{k_p}}$, где $P \subseteq \mathbb{P}$ и $k_p \in \mathbb{N}_1$ для всех $p \in P$.

Полное описание класса $\mathfrak{D}$ автору неизвестно. В настоящей работе даётся частичное описание этого класса при некоторых достаточно общих предположениях о группе $\mathcal{R}$. Чтобы сформулировать основные результаты, напомним некоторые определения из теории абелевых групп. Пусть A — абелева группа. Определение ранга группы A дано, например, в [5, т. 1, разд. 16; 6, гл. 3, разд. 4]. Для наших целей достаточно знать, что группа A имеет ранг 1, если и только если A является либо ненулевой циклической p-группой, либо квазициклической p-группой (где p — простое число), либо изоморфна некоторой ненулевой подгруппе $\mathcal{Q}$ [5, т. 1, разд. 16, упр. 6 (b); 6, гл. 3, разд. 4, упр. (2)]. Группа A называется *делимой*, если $kA = A$ для всех $k \in \mathbb{N}_1$, *p-делимой* для простого числа p, если $pA = A$, и *редуцированной*, если она не содержит ненулевых делимых подгрупп [5, т. 1, разд. 20, 21; 6, гл. 4, разд. 1]. Наконец, группа A называется *копериодической*, если A имеет прямое дополнение во всякой своей абелевой надгруппе B, такой, что B/A не имеет кручения [5, т. 1, разд. 54; гл. 9, разд. 6].

Напомним, что число $d \in \mathbb{N}_1$ называется *свободным от квадратов*, если $d = p_1 \dots p_l$, где $l \in \mathbb{N}_0$ и $p_1, \dots, p_l$ — различные простые числа; поле называется *простым*, если оно не содержит собственных подполей. Известно, что поле просто тогда и только тогда, когда оно изоморфно $\mathbb{Q}$ или $\mathbb{Z}_p$ для некоторого простого числа p. Приведём основные результаты настоящей работы (см. теорему 1):

1. Если группа $\mathcal{R}$ периодична или конечно порождена, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong \mathbb{Z}_d$ для некоторого свободного от квадратов числа $d \in \mathbb{N}_1$.
2. Если группа $\mathcal{R}$ не редуцирована, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong \cong \mathbb{Z}_d \oplus \mathbb{Q}$ для некоторого свободного от квадратов числа $d \in \mathbb{N}_1$.
3. Если группа $\mathcal{R}$ является прямой суммой подгрупп ранга 1, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong \mathbb{Z}_d$ или $R \cong \mathbb{Z}_d \oplus \mathbb{Q}$ для некоторого свободного от квадратов числа $d \in \mathbb{N}_1$ (другими словами, когда R изоморфно прямой сумме конечного числа простых полей различных характеристик).
4. Если группа $\mathcal{R}$ редуцирована и копериодична, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong \prod\limits_{p \in P} \mathbb{Z}_p$ для некоторого множества P простых чисел.

Здесь и далее через $\cong$ обозначается отношение изоморфизма. Доказательство основных результатов работы основано на том, что все кольца из класса $\mathfrak{D}$ являются E-кольцами (утверждение 1; определение E-кольца приведено в п. 3), а именно используются результаты П. Шульца [7] и Р. Боушелла и Шульца [8], в которых описаны E-кольца,

аддитивные группы которых удовлетворяют условиям, указанным в основных результатах, за исключением условия конечной порождённости (см. лемму 5).

Результаты работы позволяют описать все алгебры над полями, принадлежащие $\mathfrak{D}$, а именно: если R — алгебра над каким-либо полем F, то $R \in \mathfrak{D}$ тогда и только тогда, когда либо $R \cong F$, причём F — простое поле, либо $R = \{0\}$. Этот факт легко вывести как из приведённых выше основных, так и из некоторых промежуточных результатов (утверждение 2). В частности, поле принадлежит классу $\mathfrak{D}$, если и только если оно является простым.

1. Случай, когда R является прямым произведением

Введём некоторые обозначения. Через $\mathrm{Fun}(X, Y)$ обозначим множество всех функций из X в Y. Если $\varphi \in \mathrm{Fun}(X, Y)$ и $\psi \in \mathrm{Fun}(Y, Z)$, то $\psi(\varphi) \in \mathrm{Fun}(X, Z)$ — композиция функций φ и ψ (в этом порядке). Пусть также $\mathrm{Hom}(G, H)$ — множество всех гомоморфизмов группы G в группу H. Если H — абелева группа, то $\mathrm{Hom}(G, H)$ является абелевой группой относительно поточечного сложения. Кроме того, через $\mathrm{End}\, A$ обозначается кольцо всех эндоморфизмов абелевой группы A, в котором сложение осуществляется поточечно, а умножением является композиция.

Пусть G — группа и A — абелева группа. Удобно рассматривать множество $\mathrm{Fun}(G, A)$ как левый модуль над групповым кольцом $\mathbb{Z}G$, в котором сложение осуществляется поточечно, а действие кольца $\mathbb{Z}G$ определяется формулой $\left(\left(\sum\limits_{i=1}^{s} k_i g_i\right)\varphi\right)(x) = \sum\limits_{i=1}^{s} k_i \varphi(x g_i)$, где $\varphi \in \mathrm{Fun}(G, A)$; $x \in G$; $s \in \mathbb{N}_0$; $k_i \in \mathbb{Z}$ и $g_i \in G$ для всех $i \in \{1, \ldots, s\}$. Такой подход используется в работе [2] (там рассматривается несколько другое действие, однако его отличие от действия настоящей работы не принципиально). Очевидно, что если $\varphi \in \mathrm{Fun}(G, A)$ и $g \in G$, то

$$\partial_g \varphi = (g - 1)\varphi. \tag{1}$$

Поэтому

$$\varphi \in \mathrm{RM}_n(G, A) \iff \forall\, g_1, \ldots, g_{n+1} \in G\ ((g_1 - 1)\ldots(g_{n+1} - 1)\varphi = 0) \tag{2}$$

для произвольных $\varphi \in \mathrm{Fun}(G, A)$ и $n \in \mathbb{N}_{-1}$.

Замечание 2. Легко видеть, что $\mathrm{RM}_0(G, A)$ есть множество всех функций-констант из G в A, а $\mathrm{RM}_1(G, A)$ — множество всех функций вида $x \mapsto \eta(x) + a$ $(x \in G)$, где $\eta \in \mathrm{Hom}(G, A)$ и $a \in A$ [1, разд. 2; 2, замечание 5]. Поэтому если $\mathrm{Hom}(G, A) = \{0\}$, то $\mathrm{RM}(G, A)$ состоит из всех функций-констант из G в A, т. е. совпадает с $\mathrm{RM}_0(G, A)$.

Очевидно, что $\Pi_{-1}(R, m) = \{0\} = \mathrm{RM}_{-1}(\mathcal{R}^m, \mathcal{R})$. Используя приведённое выше описание $\mathrm{RM}_0(G, A)$, получаем, что $\Pi_0(R, m) = \mathrm{RM}_0(\mathcal{R}^m, \mathcal{R})$. Кроме того, $\Pi_n(R, 0) = \Pi_0(R, 0) = \mathrm{RM}_0(\mathcal{R}^0, \mathcal{R}) = \mathrm{RM}_{n'}(\mathcal{R}^0, \mathcal{R})$ для всех $n, n' \in \mathbb{N}_0$, так как все функции из одноэлементного множества R^0 в R являются константами.

Замечание 3. Пусть $\xi \in \mathrm{Hom}(G, H)$ для некоторой группы H. Продолжение этого гомоморфизма до гомоморфизма кольца $\mathbb{Z}G$ в кольцо $\mathbb{Z}H$ по линейности также будем обозначать через ξ. Непосредственно проверяется, что $\alpha(\psi(\xi)) = (\xi(\alpha)\psi)(\xi)$ для любых $\psi \in \mathrm{Fun}(H, A)$ и $\alpha \in \mathbb{Z}G$. Кроме того, функция $\psi \mapsto \psi(\xi)$ $(\psi \in \mathrm{Fun}(H, A))$ является гомоморфизмом аддитивной группы $\mathrm{Fun}(H, A)$ в аддитивную группу $\mathrm{Fun}(G, A)$. Таким образом, эта функция обладает свойством, аналогичным полулинейности. Легко также видеть, что если $\eta \in \mathrm{Hom}(A, B)$, где B — абелева группа, то функция $\varphi \mapsto \eta(\varphi)$ $(\varphi \in \mathrm{Fun}(G, A))$ — гомоморфизм $\mathbb{Z}G$-модуля $\mathrm{Fun}(G, A)$ в $\mathbb{Z}G$-модуль $\mathrm{Fun}(G, B)$.

Соберём нужные свойства множеств $\mathrm{RM}_n(G, A)$.

Замечание 4. Пусть $n \in \mathbb{N}_{-1}$. Тогда:

1) $\mathrm{RM}_n(G, A)$ — подмодуль $\mathbb{Z}G$-модуля $\mathrm{Fun}(G, A)$;
2) если A — подгруппа некоторой абелевой группы B, то $\mathrm{RM}_n(G, A) = \mathrm{RM}_n(G, B) \cap$ $\cap\, \mathrm{Fun}(G, A)$;
3) если $\varphi \in \mathrm{RM}_n(G, A)$, то $\varphi|_H \in \mathrm{RM}_n(H, A)$ для любой подгруппы H группы G;
4) если $\psi \in \mathrm{RM}_n(H, A)$ и $\xi \in \mathrm{Hom}(G, H)$ для некоторой группы H, то $\psi(\xi) \in$ $\in \mathrm{RM}_n(G, A)$;
5) если $\varphi \in \mathrm{RM}_n(G, A)$ и $\eta \in \mathrm{Hom}(A, B)$ для некоторой абелевой группы B, то $\eta(\varphi) \in \mathrm{RM}_n(G, B)$.

Все эти свойства проверяются непосредственно с использованием эквивалентности (2) и замечания 3 (см. также [2, замечания 2–4]).

Пусть $G = \prod\limits_{i \in I} G_i$ и $A = \prod\limits_{i \in I} A_i$, где G_i — группа и A_i — абелева группа для каждого $i \in I$. Для произвольного $i \in I$ считаем G_i подгруппой группы G, отождествляя произвольный элемент $g_i \in G_i$ с $(g_j : j \in I) \in G$, где $g_j = 1$ при всех $j \in I \setminus \{i\}$. Обозначим через ξ_i проекцию G на G_i, а через η_i — проекцию A на A_i $(i \in I)$.

Лемма 1. Предположим, что $\mathrm{Hom}(\prod_{j \in I \setminus \{i\}} G_j, A_i) = \{0\}$ для всех $i \in I$. Пусть $\varphi \in \mathrm{Fun}(G, A)$ и $n \in \mathbb{N}_{-1}$. Тогда

$$\varphi \in \mathrm{RM}_n(G, A) \iff \forall\, i \in I\ (\eta_i(\varphi) = \eta_i(\varphi(\xi_i)) \,\&\, \eta_i(\varphi|_{G_i}) \in \mathrm{RM}_n(G_i, A_i)).$$

Доказательство. Докажем импликацию «$\Longrightarrow$». Пусть $\varphi \in \mathrm{RM}_n(G, A)$ и $i \in I$. Обозначим через H_i подгруппу G, состоящую из всех $(g_j : j \in I) \in G$, таких, что $g_i = 1$. Тогда $H_i \cong \prod\limits_{j \in I \setminus \{i\}} G_j$ и G является прямым произведением подгрупп H_i и G_i. Пусть также $g_i \in G_i$. Из свойств 1, 3 и 5 замечания 4 следует, что $\eta_i((g_i\varphi)|_{H_i}) \in$ $\in \mathrm{RM}_n(H_i, A_i) \subseteq \mathrm{RM}(H_i, A_i)$. Но ввиду замечания 2 $\mathrm{RM}(H_i, A_i)$ состоит из функций-констант из $\mathrm{Fun}(H_i, A_i)$, так как $\mathrm{Hom}(H_i, A_i) = \{0\}$ согласно предположению леммы. Поэтому $\eta_i(\varphi(h_i g_i)) = \eta_i(\varphi(g_i)) = \eta_i(\varphi(\xi_i(h_i g_i)))$ для любого $h_i \in H_i$. Таким образом, $\eta_i(\varphi) = \eta_i(\varphi(\xi_i))$. Кроме того, из свойств 3 и 5 замечания 4 вытекает, что $\eta_i(\varphi|_{G_i}) \in$ $\in \mathrm{RM}_n(G_i, A_i)$. Импликация «$\Longrightarrow$» доказана.

Докажем теперь импликацию «$\Longleftarrow$». Предположим, что $\eta_i(\varphi) = \eta_i(\varphi(\xi_i))$ и $\eta_i(\varphi|_{G_i}) \in \mathrm{RM}_n(G_i, A_i)$ для всех $i \in I$. Пусть $g_1, \ldots, g_{n+1} \in G$. Положим для краткости $\alpha = (g_1 - 1) \ldots (g_{n+1} - 1)$. Тогда для каждого $i \in I$

$$\eta_i(\alpha\varphi) = \alpha(\eta_i(\varphi)) = \alpha(\eta_i(\varphi(\xi_i))) = ((\xi_i(g_1) - 1) \ldots (\xi_i(g_{n+1}) - 1)(\eta_i(\varphi)))(\xi_i) = 0$$

ввиду замечания 3 и эквивалентности (2). Следовательно, $\alpha\varphi = 0$. Таким образом, $\varphi \in \mathrm{RM}_n(G, A)$ (см. эквивалентность (2)). Импликация «$\Longleftarrow$» доказана. ■

Лемма 1 может быть сформулирована следующим образом: если $\mathrm{Hom}\left(\prod\limits_{j \in I \setminus \{i\}} G_j, A_i\right) =$ $= \{0\}$ для всех $i \in I$, то $\mathrm{RM}_n(G, A)$ $(n \in \mathbb{N}_{-1})$ состоит из тех и только тех функций $\varphi \in \mathrm{Fun}(G, A)$, которые имеют вид

$$(x_i : i \in I) \mapsto (\varphi_i(x_i) : i \in I) \quad (x_i \in G_i),$$

где $\varphi_i \in \mathrm{RM}_n(G_i, A_i)$ при любом $i \in I$. Очевидно, что функции φ_i определяются такой функцией φ однозначно, а именно: $\varphi_i = \eta_i(\varphi|_{G_i})$ для всех $i \in I$.

Пусть теперь $R = \prod\limits_{i\in I} R_i$, где R_i — ассоциативное коммутативное кольцо для каждого $i \in I$. Тогда R^m естественно отождествляется с $\prod\limits_{i\in I} R_i^m$. Для произвольного $i \in I$ считаем R_i идеалом кольца R, отождествляя произвольный элемент $r_i \in R_i$ с $(r_j : j \in I) \in R$, где $r_j = 0$ при всех $j \in I \setminus \{i\}$. Обозначим через η_i проекцию R на R_i, а через ξ_i — проекцию R^m на R_i^m, определённую равенством $\xi_i(x_1, \ldots, x_m) = (\eta_i(x_1), \ldots, \eta_i(x_m))$ для всех $x_1, \ldots, x_m \in R$.

Замечание 5. Пусть $\varphi \in \mathrm{Fun}(R^m, R)$ и $n \in \mathbb{N}_{-1}$. Тогда непосредственно проверяется, что

$$\varphi \in \Pi_n(R, m) \iff \forall\, i \in I\ (\eta_i(\varphi) = \eta_i(\varphi(\xi_i)) \,\&\, \eta_i(\varphi|_{R_i^m}) \in \Pi_n(R_i, m)).$$

Лемма 2.

1. Пусть $n \in \mathbb{N}_{-1}$. Тогда если $\Pi_n(R, m) = \mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$, то $\Pi_n(R_i, m) = \mathrm{RM}_n(\mathcal{R}_i^m, \mathcal{R}_i)$ для всех $i \in I$.
2. Если $\Pi(R, m) = \mathrm{RM}(\mathcal{R}^m, \mathcal{R})$, то $\Pi(R_i, m) = \mathrm{RM}(\mathcal{R}_i^m, \mathcal{R}_i)$ для всех $i \in I$.

Доказательство. Докажем утверждение 1. Предположим, что $\Pi_n(R, m) = \mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$. Пусть $i \in I$ и $\varphi_i \in \mathrm{RM}_n(\mathcal{R}_i^m, \mathcal{R}_i)$. Положим $\varphi = \varphi_i(\xi_i)$. Тогда $\varphi \in \mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$ ввиду свойств 4 и 2 замечания 4. Следовательно, $\varphi \in \Pi_n(R, m)$. Из замечания 5 вытекает, что $\varphi_i = \eta_i(\varphi_i) = \eta_i(\varphi|_{R_i^m}) \in \Pi_n(R_i, m)$ (очевидно, что $\varphi_i = \varphi|_{R_i^m}$). Таким образом, $\mathrm{RM}_n(\mathcal{R}_i^m, \mathcal{R}_i) \subseteq \Pi_n(R_i, m)$, а обратное включение имеет место ввиду замечания 1. Утверждение 1 доказано.

Утверждение 2 доказывается так же, за исключением того, что $\varphi \in \Pi_{n'}(R, m)$ и $\varphi_i \in \Pi_{n'}(R_i, m)$ для некоторого $n' \in \mathbb{N}_{-1}$, вообще говоря, отличного от n. Поэтому если $\Pi(R, m) = \mathrm{RM}(\mathcal{R}^m, \mathcal{R})$, то $\mathrm{RM}(\mathcal{R}_i^m, \mathcal{R}_i) \subseteq \Pi(R_i, m)$ для любого $i \in I$, а обратное включение следует из замечания 1. ■

Лемма 3. Предположим, что $\mathrm{Hom}(\prod_{j\in I\setminus\{i\}} \mathcal{R}_j, \mathcal{R}_i) = \{0\}$ для всех $i \in I$. Пусть $n \in \mathbb{N}_{-1}$. Тогда

$$\Pi_n(R, m) = \mathrm{RM}_n(\mathcal{R}^m, \mathcal{R}) \iff \forall\, i \in I\ (\Pi_n(R_i, m) = \mathrm{RM}_n(\mathcal{R}_i^m, \mathcal{R}_i)).$$

В частности, $R \in \mathfrak{D}$, если и только если $R_i \in \mathfrak{D}$ для каждого $i \in I$.

Доказательство. Ввиду утверждения 1 леммы 2 достаточно доказать лишь импликацию «$\Longleftarrow$». Пусть $\varphi \in \mathrm{Fun}(R^m, R)$. Из предположения леммы следует, что $\mathrm{Hom}\left(\prod\limits_{j\in I\setminus\{i\}} \mathcal{R}_j^m, \mathcal{R}_i^m\right) = \{0\}$ для всех $i \in I$. Поэтому

$$\varphi \in \mathrm{RM}_n(\mathcal{R}^m, \mathcal{R}) \iff \forall\, i \in I\ (\eta_i(\varphi) = \eta_i(\varphi(\xi_i)) \,\&\, \eta_i(\varphi|_{R_i^m}) \in \mathrm{RM}_n(\mathcal{R}_i^m, \mathcal{R}_i))$$

согласно лемме 1. Требуемая импликация «$\Longleftarrow$» непосредственно вытекает из замечания 5 и этой эквивалентности. ■

Отметим, что если множество I конечно (в этом случае $R = \bigoplus\limits_{i\in I} R_i$), то предположение леммы 3 выполнено тогда и только тогда, когда $\mathrm{Hom}(\mathcal{R}_i, \mathcal{R}_j) = \{0\}$ для любых различных $i, j \in I$.

2. Случай, когда R — простое поле

Далее предполагаем, что кольцо R содержит единицу. Для произвольного $k \in \mathbb{Z}$ будем обозначать элемент $k1$ этого кольца через k, если это не вызывает недоразумений. Аналогично предыдущему, рассматриваем множество $\mathrm{Fun}(\mathcal{R}^m, R)$ как модуль над групповым кольцом $R\mathcal{R}^m$. Для удобства записи элементов этого группового кольца введём формальный символ $\mathbf{g}$ и вместо произвольного элемента $v \in \mathcal{R}^m$ будем писать $\mathbf{g}^v$. При этом, разумеется, $\mathbf{g}^{u+v} = \mathbf{g}^u\mathbf{g}^v$ для любых $u, v \in \mathcal{R}^m$; обозначим также $\mathbf{g}^0$ через 1. Таким образом, элементы группового кольца $R\mathcal{R}^m$ будут записываться в виде $\sum\limits_{i=1}^{s} r_i\mathbf{g}^{v_i}$, где $s \in \mathbb{N}_0$, $r_i \in R$ и $v_i \in \mathcal{R}^m$ для всех $i \in \{1, \ldots, s\}$; результатом действия этого элемента на произвольную функцию $\varphi \in \mathrm{Fun}(\mathcal{R}^m, R)$ является функция $x \mapsto \sum\limits_{i=1}^{s} r_i\varphi(x + v_i)$ $(x \in \mathcal{R}^m)$. Сложение в $\mathrm{Fun}(\mathcal{R}^m, R)$ осуществляется поточечно. Формула (1) для производной в данной ситуации приобретает вид

$$\partial_v\varphi = (\mathbf{g}^v - 1)\varphi, \tag{3}$$

где $\varphi \in \mathrm{Fun}(\mathcal{R}^m, R)$ и $v \in \mathcal{R}^m$. Отметим, что если R конечно, то $\mathrm{Fun}(\mathcal{R}^m, R)$ и $R\mathcal{R}^m$ изоморфны как $R\mathcal{R}^m$-модули; изоморфизмом первого из этих модулей на второй является функция $\varphi \mapsto \sum\limits_{x\in\mathcal{R}^m} \varphi(x)\mathbf{g}^{-x}$ $(\varphi \in \mathrm{Fun}(\mathcal{R}^m, R))$.

Очевидно, что для любого $n \in \mathbb{N}_{-1}$ множества $\Pi_n(R, m)$ и $\mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$ являются подмодулями $R\mathcal{R}^m$-модуля $\mathrm{Fun}(\mathcal{R}^m, R)$. Легко также видеть, что для любого $v \in \mathcal{R}^m$ оператор взятия производной по направлению v (обозначаемый через ∂_v) является эндоморфизмом $R\mathcal{R}^m$-модуля $\mathrm{Fun}(\mathcal{R}^m, R)$, отображающим $\mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$ в $\mathrm{RM}_{n-1}(\mathcal{R}^m, \mathcal{R})$ для каждого $n \in \mathbb{N}_0$.

Для каждого $i \in \{1, \ldots, m\}$ положим $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, где единица стоит на i-м месте. В зависимости от контекста e_i может быть элементом как $\mathcal{R}^m$, так и $\mathbb{N}_0^m$. Будем рассматривать $\mathbb{N}_0^m$ как прямое произведение m упорядоченных множеств $\mathbb{N}_0$ с обычным отношением порядка $\leqslant$. Другими словами, $k \leqslant l$ (или $l \geqslant k$), где $k = (k_1, \ldots, k_m)$ и $l = (l_1, \ldots, l_m)$ — произвольные наборы из $\mathbb{N}_0^m$, если и только если $k_i \leqslant l_i$ для всех $i \in \{1, \ldots, m\}$. Для произвольной функции $\varphi \in \mathrm{Fun}(\mathcal{R}^m, R)$ и произвольного $k = (k_1, \ldots, k_m) \in \mathbb{N}_0^m$ положим $\partial_e^k\varphi = \partial_{e_1}^{k_1} \ldots \partial_{e_m}^{k_m}\varphi$.

Пусть теперь R — поле, c — его характеристика. Положим $I_c = \mathbb{N}_0$, если $c = 0$, и $I_c = \{0, \ldots, c - 1\}$ в противном случае. Пусть также

$$I_c^m(n) = \{(k_1, \ldots, k_m) \in I_c^m : k_1 + \cdots + k_m \leqslant n\}$$

для произвольного $n \in \mathbb{N}_{-1}$. Как обычно, если $y \in R$ и $t \in I_c$, то

$$\binom{y}{t} = \frac{y(y-1)\ldots(y-t+1)}{t!}.$$

Кроме того, для произвольных $x = (x_1, \ldots, x_m) \in R^m$ и $k = (k_1, \ldots, k_m) \in I_c^m$ положим

$$\binom{x}{k} = \binom{x_1}{k_1} \ldots \binom{x_m}{k_m}.$$

Лемма 4. Предположим, что R — простое поле. Пусть $\varphi \in \mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$, где $n \in \mathbb{N}_{-1}$. Тогда для любого $x \in R^m$ имеет место равенство

$$\varphi(x) = \sum_{k \in I_c^m(n)} \binom{x}{k} (\partial_e^k \varphi)(0), \tag{4}$$

где c — характеристика поля R.

Доказательство. Пусть $l = (l_1, \ldots, l_m) \in I_c^m$. Непосредственно проверяется, что

$$\begin{gathered}\mathbf{g}^l \varphi = \left(\prod_{i=1}^m (\mathbf{g}^{e_i})^{l_i} \right) \varphi = \left(\prod_{i=1}^m ((\mathbf{g}^{e_i} - 1) + 1)^{l_i} \right) \varphi = \left(\prod_{i=1}^m \left(\sum_{k_i=0}^{l_i} \binom{l_i}{k_i} (\mathbf{g}^{e_i} - 1)^{k_i} \right) \right) \varphi = \\ = \sum_{k \in I_c^m : k \leqslant l} \binom{l}{k} \partial_e^k \varphi = \sum_{k \in I_c^m(n) : k \leqslant l} \binom{l}{k} \partial_e^k \varphi = \sum_{k \in I_c^m(n)} \binom{l}{k} \partial_e^k \varphi.\end{gathered} \tag{5}$$

Здесь мы воспользовались формулой (3) и следующими фактами:

— если $k \in I_c^m \setminus I_c^m(n)$, то $\partial_e^k \varphi = 0$ (вытекает из того, что $\varphi \in \mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$);

— если $k \in I_c^m$ и $k \not\leqslant l$, то $\binom{l}{k} = 0$ (вытекает из того, что $\binom{y}{t} = 0$ при любых $t \in I_c$ и $y \in \{0, \ldots, t-1\}$).

Взяв значения функций из (5) в 0, получаем

$$\varphi(l) = \sum_{k \in I_c^m(n)} \binom{l}{k} (\partial_e^k \varphi)(0). \tag{6}$$

Из этой формулы непосредственно вытекает утверждение леммы в случае, когда $c \neq 0$.

С этого момента и до окончания доказательства леммы будем рассматривать случай, когда $c = 0$. Тогда можно считать, что $R = \mathbb{Q}$. Для доказательства леммы в этом случае воспользуемся индукцией по n. При $n = -1$ утверждение леммы верно, так как $\mathrm{RM}_{-1}(\mathcal{Q}^m, \mathcal{Q}) = \{0\}$ и $I_0^m(-1) = \varnothing$. Пусть теперь $n \in \mathbb{N}_0$ и формула (4) верна для всех функций из $\mathrm{RM}_{n-1}(\mathcal{Q}^m, \mathcal{Q})$ и всех $x \in \mathbb{Q}^m$. Определим функцию $\pi \colon \mathbb{Q}^m \to \mathbb{Q}$ равенством

$$\pi(x) = \sum_{l \in I_0^m(n)} \binom{x}{l} (\partial_e^l \varphi)(0)$$

для всех $x \in \mathbb{Q}^m$ и положим $\psi = \varphi - \pi$. Очевидно, что $\pi \in \Pi_n(\mathbb{Q}, m) \subseteq \mathrm{RM}_n(\mathcal{Q}^m, \mathcal{Q})$ (см. замечание 1) и, следовательно, $\psi \in \mathrm{RM}_n(\mathcal{Q}^m, \mathcal{Q})$. Для завершения доказательства требуется показать, что $\psi = 0$. Так как $\psi(x) = 0$ для всех $x \in I_0^m = \mathbb{N}_0^m$ (ввиду формулы (6)), для этого достаточно доказать, что ψ — константа. В свою очередь, последнее условие эквивалентно тому, что $\partial_v \psi = 0$ для всех $v \in \mathcal{Q}^m$ (см. замечание 2). Положим $L = \{v \in \mathcal{Q}^m : \partial_v \psi = 0\}$ и покажем, что $L = \mathcal{Q}^m$. Это вытекает из следующих утверждений:

1) $e_1, \ldots, e_m \in L$;
2) L — подгруппа $\mathcal{Q}^m$;
3) если $v \in L$ и $d \in \mathbb{N}_1$, то $v/d \in L$.

Докажем утверждение 1. Пусть $i \in \{1, \ldots, m\}$ и $x \in \mathbb{Q}^n$. Из известного и легко проверяемого равенства

$$\binom{y+1}{t} - \binom{y}{t} = \binom{y}{t-1},$$

справедливого для всех $y \in \mathbb{Q}$ и $t \in \mathbb{N}_1$, следует, что

$$(\partial_{e_i} \pi)(x) = \sum_{l \in I_0^m(n) : l \geqslant e_i} \binom{x}{l - e_i} (\partial_e^l \varphi)(0) = \sum_{k \in I_0^m(n-1)} \binom{x}{k} (\partial_e^{k+e_i} \varphi)(0).$$

Здесь мы воспользовались тем, что $l \mapsto l - e_i$ — биекция $\{l \in I_0^m(n) : l \geqslant e_i\}$ на $I_0^m(n-1)$. С другой стороны, предположение индукции, применённое к $\partial_{e_i}\varphi$, даёт равенство

$$(\partial_{e_i}\varphi)(x) = \sum_{k \in I_0^m(n-1)} \binom{x}{k} (\partial_e^{k+e_i}\varphi)(0)$$

(очевидно, что операторы взятия производных по разным направлениям коммутируют). Поэтому $\partial_{e_i}\psi = \partial_{e_i}\varphi - \partial_{e_i}\pi = 0$, т. е. $e_i \in L$. Утверждение 1 доказано.

Утверждение 2 непосредственно вытекает из равенств $\partial_0\psi = 0$, $\partial_{u+v}\psi = \mathbf{g}^u\partial_v\psi + \partial_u\psi$ и $\partial_{-v}\psi = -\mathbf{g}^{-v}\partial_v\psi$, имеющих место для любых $u, v \in \mathcal{Q}^m$. Эти равенства легко получить, используя формулу (3).

Докажем утверждение 3. Обозначим через Δ фундаментальный идеал групповой алгебры $\mathbb{Q}\mathcal{Q}^m$, т. е. ядро гомоморфизма $\mathbb{Q}$-алгебр $\tau\colon \mathbb{Q}\mathcal{Q}^m \to \mathbb{Q}$, такого, что $\tau\left(\sum_{i=1}^{s} r_i\mathbf{g}^{v_i}\right) = \sum_{i=1}^{s} r_i$ для всех $s \in \mathbb{N}_0$, $r_i \in \mathbb{Q}$ и $v_i \in \mathcal{Q}^m$ ($i \in \{1, \ldots, s\}$). Легко видеть, что Δ порождается как векторное пространство над $\mathbb{Q}$ и как $\mathbb{Q}\mathcal{Q}^m$-идеал множеством $\{\mathbf{g}^v - 1 : v \in \mathcal{Q}^m\}$.

Пусть $v \in L$ и $d \in \mathbb{N}_1$. Положим $u = v/d$. Обозначим через $\Delta^{(n)}$ n-ю степень идеала Δ; при $n = 0$ полагаем $\Delta^{(0)} = \mathbb{Q}\mathcal{Q}^m$. Так как $\partial_u\psi \in \mathrm{RM}_{n-1}(\mathcal{Q}^m, \mathcal{Q})$, из формулы (3) вытекает, что $\Delta^{(n)}\partial_u\psi = \{0\}$ (см. также лемму 2 из [2]). Непосредственно проверяется (с использованием формулы (3)), что $\partial_v\psi = \alpha\partial_u\psi$, где $\alpha = 1 + \mathbf{g}^u + \cdots + \mathbf{g}^{(d-1)u} \in \mathbb{Q}\mathcal{Q}^m$. Теперь воспользуемся тем, что по модулю $\Delta^{(n)}$ любой элемент $\mathbb{Q}\mathcal{Q}^m \setminus \Delta$ (в частности, α) обратим, а именно: пусть $\delta = 1 - \alpha/d$ и $\beta = (1 + \delta + \cdots + \delta^{n-1})/d$. Тогда легко видеть, что $\delta \in \Delta$ (так как $\tau(\alpha) = d$) и $\beta\alpha = 1 - \delta^n \equiv 1 \pmod{\Delta^{(n)}}$. Поэтому $\partial_u\psi = \beta\alpha\partial_u\psi = \beta\partial_v\psi = 0$, т. е. $u = v/d \in L$. Утверждение 3 доказано. ■

Формулу (4) можно рассматривать как вариант формулы Ньютона. Другие варианты формулы Ньютона см., например, в [4, разд. 4; 9, разд. 6.2]. Из доказательства леммы 4 (см. формулу (6)) вытекает, что формула (4) верна, если R — произвольное поле характеристики 0 и $\varphi \in \Pi_n(R, m)$. (Здесь используется также известное утверждение о том, что если R — поле, X — его бесконечное подмножество, а φ и π — функции из $\Pi(R, m)$ такие, что $\varphi|_{X^m} = \pi|_{X^m}$, то $\varphi = \pi$.) Вероятно, этот факт известен специалистам, так же как и формула (4) для функций из $\Pi_n(\mathbb{Z}_p, m)$, где p — простое число. Однако нашей задачей является доказательство формулы (4) для функций из $\mathrm{RM}_n(\mathcal{R}^m, \mathcal{R})$ (а не просто из $\Pi_n(R, m)$), если R — простое поле. Это позволяет показать, что $\mathrm{RM}_n(\mathcal{R}^m, \mathcal{R}) \subseteq \Pi_n(R, m)$ для всех $n \in \mathbb{N}_{-1}$, так как в правой части формулы (4) стоит $\pi(x)$ для некоторой функции $\pi \in \Pi_n(R, m)$. Обратное включение имеет место ввиду замечания 1. Таким образом, получаем

Следствие 1. Все простые поля принадлежат классу $\mathfrak{D}$.

Для конечных простых полей это утверждение известно (см. [3, свойство B7], где оно сформулировано на языке степеней). При $p = 2$ оно отмечается в [1, разд. 2].

3. Доказательство основных результатов

Ассоциативное кольцо S с единицей называется *E-кольцом*, если выполнено любое из следующих эквивалентных условий:

- всякий эндоморфизм группы $\mathcal{S}$ является эндоморфизмом S как правого S-модуля, т. е. имеет вид $x \mapsto sx$ ($x \in S$) при некотором $s \in S$;
- кольцо S коммутативно и $S \cong \mathrm{End}\,\mathcal{S}$;
- кольцо $\mathrm{End}\,\mathcal{S}$ коммутативно

(см. [6, гл. 18, разд. 6; 7, разд. 3; 8, разд. 1]). Понятие E-кольца понадобится потому, что всякое кольцо из класса $\mathfrak{D}$ является E-кольцом (см. утверждение 1 далее).

Пусть $P \subseteq \mathbb{P}$. Через $\mathbb{Z}[1/P]$ будем обозначать подкольцо $\mathbb{Q}$, порождённое множеством $\{1\} \cup \{1/p \,|\, p \in P\}$. Другими словами, $\mathbb{Z}[1/P]$ — это множество всех k/l, где $k \in \mathbb{Z}$, $l \in \mathbb{N}_1$, причём все простые делители l принадлежат P. В частности, $\mathbb{Z}[1/\varnothing] = \mathbb{Z}$ и $\mathbb{Z}[1/\mathbb{P}] = \mathbb{Q}$. Для произвольного простого числа p через $\mathbb{J}_p$ будем обозначать кольцо целых p-адических чисел.

Замечание 6. Пусть S — кольцо с единицей, аддитивная группа которого является группой ранга 1 без кручения. Тогда для любого $s \in S$ существуют $k_s \in \mathbb{Z}$ и $l_s \in \mathbb{N}_1$, удовлетворяющие равенству $l_s s = k_s 1$. Непосредственно проверяется, что $s \mapsto k_s/l_s$ $(s \in S)$ — корректно определённый инъективный гомоморфизм из кольца S в поле $\mathbb{Q}$. Образ S при этом гомоморфизме содержит $\mathbb{Z}$, поэтому этот образ есть $\mathbb{Z}[1/X]$, где $X = \{p \in \mathbb{P} : 1 \in pS\}$ (см. теорему 2.2 из работы [10] и её доказательство). Таким образом, $S \cong \mathbb{Z}[1/X]$.

Утверждение 1. Кольцо R является E-кольцом (в частности, имеет единицу) тогда и только тогда, когда $\Pi_1(R, 1) = \mathrm{RM}_1(\mathcal{R}, \mathcal{R})$. В частности, если $R \in \mathfrak{D}$, то R — E-кольцо.

Доказательство. Согласно замечанию 2, $\mathrm{RM}_1(\mathcal{R}, \mathcal{R})$ есть множество всех функций вида $x \mapsto \epsilon(x) + r$ $(x \in \mathcal{R})$, где $\epsilon \in \operatorname{End}\mathcal{R}$ и $r \in \mathcal{R}$. Поэтому если R — E-кольцо, то $\mathrm{RM}_1(\mathcal{R}, \mathcal{R}) \subseteq \Pi_1(R, 1)$ и, следовательно, $\Pi_1(R, 1) = \mathrm{RM}_1(\mathcal{R}, \mathcal{R})$ ввиду замечания 1. Обратно, пусть $\Pi_1(R, 1) = \mathrm{RM}_1(\mathcal{R}, \mathcal{R})$. Тогда из описания множества $\mathrm{RM}_1(\mathcal{R}, \mathcal{R})$ вытекает, что все эндоморфизмы группы $\mathcal{R}$ имеют вид $x \mapsto sx$ $(x \in \mathcal{R})$, где $s \in R$. В частности, кольцо R имеет единицу. Таким образом, R является E-кольцом. ■

Приведём нужные свойства E-колец.

Лемма 5. Предположим, что R является E-кольцом. Тогда:

1) если $\mathcal{R} = A \oplus B$ для некоторых подгрупп A и B группы $\mathcal{R}$, то $\operatorname{Hom}(A, B) = \{0\}$;
2) если S — ассоциативное кольцо с единицей, такое, что $\mathcal{S} \cong \mathcal{R}$, то $S \cong R$;
3) если группа $\mathcal{R}$ периодична, то $R \cong \mathbb{Z}_d$ для некоторого $d \in \mathbb{N}_1$;
4) если группа $\mathcal{R}$ не редуцирована, то $R \cong \mathbb{Z}_d \oplus \mathbb{Q}$ для некоторого $d \in \mathbb{N}_1$;
5) если группа $\mathcal{R}$ является прямой суммой подгрупп ранга 1, то $R \cong \mathbb{Z}_d \oplus \oplus \bigoplus_{i=1}^{k} \mathbb{Z}[1/X_i]$, где $d \in \mathbb{N}_1$; $k \in \mathbb{N}_0$; $X_1, \ldots, X_k$ — множества простых чисел, содержащие все простые делители числа d, причём $X_i \not\subseteq X_j$ при $i \neq j$ $(i, j \in \in \{1, \ldots, k\})$;
6) если группа $\mathcal{R}$ редуцирована и копериодична, то $R \cong \prod_{p \in P} R_p$, где $P \subseteq \mathbb{P}$ и $R_p \in \{\mathbb{Z}_{p^k} : k \in \mathbb{N}_1\} \cup \{\mathbb{J}_p\}$ для каждого $p \in P$.

Доказательство. Утверждение 1 следует из свойства D E-колец [6, гл. 18, разд. 6]. Впрочем, это утверждение легко доказать непосредственно: если $\mathcal{R} = A \oplus B$, ξ — проекция $\mathcal{R}$ на A и η — ненулевой гомоморфизм из A в B, то эндоморфизмы ξ и $\eta(\xi)$ группы $\mathcal{R}$ не коммутируют. Утверждение 2 вытекает из следствия 4 в [7]. Оно тоже легко доказывается: если $\mathcal{S} \cong \mathcal{R}$, то S — E-кольцо (так как $\operatorname{End}\mathcal{S}$ изоморфно коммутативному кольцу $\operatorname{End}\mathcal{R}$) и $S \cong \operatorname{End}\mathcal{S} \cong \operatorname{End}\mathcal{R} \cong R$. Утверждение 3 следует из теоремы 1 в [7] (см. также [6, гл. 18, лемма 6.4, (i); 8, предложение 1.4, (ii)]). Утверждение 4 вытекает из теоремы 3 работы [7] (см. также [6, гл. 18, лемма 6.4, (ii); 8, предложение 1.4, (iii)]). Пусть теперь $\mathcal{R}$ является прямой суммой подгрупп ранга 1. Тогда

из теоремы 2 в [7] следует, что $R \cong \mathbb{Z}_d \oplus \bigoplus\limits_{i=1}^{k} R_i$, где $d \in \mathbb{N}_1$; $k \in \mathbb{N}_0$; $R_1, \ldots, R_k$ — ассоциативные коммутативные кольца с единицей, аддитивные группы которых не имеют кручения, имеют ранг 1, являются p-делимыми для любого простого делителя p числа d и имеют попарно несравнимые типы. О типах абелевых групп без кручения ранга 1 см. [5, т. 2, разд. 85; 6, гл. 12, разд. 1]. Замечание 6 показывает, что $R_i \cong \mathbb{Z}[1/X_i]$ для каждого $i \in \{1, \ldots, k\}$, где $X_i \subseteq \mathbb{P}$. При этом множество X_i для каждого $i \in \{1, \ldots, k\}$ содержит все простые делители числа d, так как группа $\mathcal{Z}[1/X]$ p-делима (где $p \in \mathbb{P}$ и $X \subseteq \mathbb{P}$) тогда и только тогда, когда $p \in X$. Кроме того, тип $\mathcal{Z}[1/X]$ не превосходит типа $\mathcal{Z}[1/Y]$, если и только если $X \subseteq Y$ ($X, Y \subseteq \mathbb{P}$). Это показывает, что $X_i \nsubseteq X_j$ при $i \neq j$ ($i, j \in \{1, \ldots, k\}$). Таким образом, утверждение 5 доказано. Наконец, утверждение 6 доказано в теореме 3.3 в [8] (см. также [6, гл. 18, теорема 6.6]). ■

Утверждение 2. Пусть R — алгебра над каким-либо полем F. Тогда следующие условия эквивалентны:

1) $R \in \mathfrak{D}$;
2) $\Pi_1(R, 1) = \mathrm{RM}_1(\mathcal{R}, \mathcal{R})$;
3) R — E-кольцо;
4) либо $R \cong F$, причём F — простое поле, либо $R = \{0\}$.

Доказательство. Импликация $1 \Longrightarrow 2$ тривиальна, импликация $2 \Longrightarrow 3$ доказана в утверждении 1, а импликация $4 \Longrightarrow 1$ вытекает из следствия 1. Осталось доказать импликацию $3 \Longrightarrow 4$. Достаточно рассмотреть случай, когда R — ненулевое E-кольцо. Тогда группа $\mathcal{R}$ изоморфна прямой сумме $\dim_F R$ экземпляров группы $\mathcal{F}$. Из утверждения 1 леммы 5 следует, что $\dim_F R = 1$. Поэтому $\mathcal{R} \cong \mathcal{F}$ и $R \cong F$ (см. утверждение 2 леммы 5). Пусть теперь K — простое подполе поля F. Рассуждая аналогично, получим, что $F = K$. Импликация $3 \Longrightarrow 4$ доказана. ■

Замечание 7. Пусть $\pi \in \Pi(R, m)$. Пусть также I — идеал кольца R и $x_1, y_1, \ldots, x_m, y_m \in R$. Тогда легко видеть, что

$$\forall\, i \in \{1, \ldots, m\}\ (x_i \equiv y_i \pmod{I}) \implies \pi(x_1, \ldots, x_m) \equiv \pi(y_1, \ldots, y_m) \pmod{I}.$$

Лемма 6. Пусть p — простое число и $k \in \mathbb{N}_2$. Тогда $\Pi(\mathbb{Z}_{p^k}, 1) \neq \mathrm{RM}(\mathcal{Z}_{p^k}, \mathcal{Z}_{p^k})$. В частности, $\mathbb{Z}_{p^k} \notin \mathfrak{D}$.

Доказательство. Согласно теореме 2 (или теореме 1) из [2], $\mathrm{RM}(\mathcal{Z}_{p^k}, \mathcal{Z}_{p^k})$ совпадает с множеством всех функций из $\mathbb{Z}_{p^k}$ в $\mathbb{Z}_{p^k}$. Известно, что не все такие функции принадлежат $\Pi(\mathbb{Z}_{p^k}, 1)$. Действительно, из замечания 7 следует, что если функция $\varphi\colon \mathbb{Z}_{p^k} \to \mathbb{Z}_{p^k}$ удовлетворяет условиям $\varphi(0) = 0$ и $\varphi(p) = 1$, то $\varphi \notin \Pi(\mathbb{Z}_{p^k}, 1)$. ■

Лемма 7. Пусть X — собственное подмножество множества $\mathbb{P}$. Тогда $\mathrm{RM}_p(\mathcal{Z}[1/X], \mathcal{Z}[1/X]) \nsubseteq \Pi(\mathbb{Z}[1/X], 1)$ для любого $p \in \mathbb{P} \setminus X$. В частности, $\Pi(\mathbb{Z}[1/X], 1) \neq \mathrm{RM}(\mathcal{Z}[1/X], \mathcal{Z}[1/X])$ и $\mathbb{Z}[1/X] \notin \mathfrak{D}$.

Доказательство. Пусть $p \in \mathbb{P} \setminus X$. Определим функцию $\pi \in \Pi_p(\mathbb{Z}[1/X], 1)$ формулой $\pi(x) = x^p - x$ ($x \in \mathbb{Z}[1/X]$). Легко видеть, что $\mathbb{Z}[1/X]/p\mathbb{Z}[1/X] \cong \mathbb{Z}_p$ (так как каждый элемент $\mathbb{Z}[1/X]$ представим дробью, знаменатель которой обратим по модулю p). Поэтому $\pi(\mathbb{Z}[1/X]) \subseteq p\mathbb{Z}[1/X]$ и π/p отображает $\mathbb{Z}[1/X]$ в $\mathbb{Z}[1/X]$. Кроме того, $\partial_{v_1} \ldots \partial_{v_{p+1}}(\pi/p) = (\partial_{v_1} \ldots \partial_{v_{p+1}}\pi)/p = 0$ для любых $v_1, \ldots, v_{p+1} \in \mathcal{Z}[1/X]$, так как $\pi \in \mathrm{RM}_p(\mathcal{Z}[1/X], \mathcal{Z}[1/X])$ (ввиду замечания 1). Поэтому $\pi/p \in \mathrm{RM}_p(\mathcal{Z}[1/X], \mathcal{Z}[1/X])$. Но легко видеть, что $\pi/p \notin \Pi(\mathbb{Z}[1/X], 1)$. Действительно, в противном случае функ-

ция π на бесконечном подмножестве $\mathbb{Z}[1/X]$ поля $\mathbb{Q}$ была бы представима многочленом с коэффициентами из $p\mathbb{Z}[1/X]$, что невозможно. ■

Замечание 8. В работе [11] Р. Димитрич ввёл понятие коузкой абелевой группы, а именно: абелева группа A *коузка*, если и только если $\mathrm{Hom}(A, \mathcal{Z}) = \{0\}$ или, что эквивалентно, A не содержит бесконечной циклической подгруппы, имеющей в A прямое дополнение (см. теорему 3 указанной работы). Предположим, что R — E-кольцо и $\Pi(R, 1) = \mathrm{RM}(\mathcal{R}, \mathcal{R})$ (ввиду утверждения 1 это верно, если $R \in \mathfrak{D}$). Покажем, что тогда группа $\mathcal{R}$ коузка. Пусть это не так. Тогда $\mathcal{R} \cong \mathcal{Z} \oplus B$ для некоторой абелевой группы B. Утверждение 1 леммы 5 показывает, что $\mathrm{Hom}(\mathcal{Z}, B) = \{0\}$. Следовательно, $B = \{0\}$, так как для любого $b \in B$ существует гомоморфизм из $\mathcal{Z}$ в B, переводящий 1 в b. Таким образом, $\mathcal{R} \cong \mathcal{Z}$ и $R \cong \mathbb{Z}$ (см. утверждение 2 леммы 5). Но тогда $\mathrm{RM}_p(\mathcal{R}, \mathcal{R}) \nsubseteq \Pi(R, 1)$ для любого $p \in \mathbb{P}$ ввиду леммы 7 (так как $\mathbb{Z} = \mathbb{Z}[1/\varnothing]$), что противоречит сделанному выше предположению.

Лемма 8. Пусть p — простое число. Тогда $\mathrm{RM}_p(\mathcal{J}_p, \mathcal{J}_p) \nsubseteq \Pi(\mathbb{J}_p, 1)$. В частности, $\Pi(\mathbb{J}_p, 1) \neq \mathrm{RM}(\mathcal{J}_p, \mathcal{J}_p)$ и $\mathbb{J}_p \notin \mathfrak{D}$.

Доказательство. Известно, что $\mathbb{J}_p/p\mathbb{J}_p \cong \mathbb{Z}_p$. Кроме того, $\mathbb{J}_p$ является бесконечным подкольцом поля p-адических чисел. Поэтому, рассуждая так же, как в доказательстве леммы 7, легко показать, что функция $x \mapsto (x^p - x)/p$ $(x \in \mathbb{J}_p)$ принадлежит $\mathrm{RM}_p(\mathcal{J}_p, \mathcal{J}_p)$, но не $\Pi(\mathbb{J}_p, 1)$. ■

Теорема 1.

1. Если группа $\mathcal{R}$ периодична или конечно порождена, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong \mathbb{Z}_d$ для некоторого свободного от квадратов числа $d \in \mathbb{N}_1$.
2. Если группа $\mathcal{R}$ не редуцирована, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong \mathbb{Z}_d \oplus \mathbb{Q}$ для некоторого свободного от квадратов числа $d \in \mathbb{N}_1$.
3. Если группа $\mathcal{R}$ является прямой суммой подгрупп ранга 1, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong \mathbb{Z}_d$ или $R \cong \mathbb{Z}_d \oplus \mathbb{Q}$ для некоторого свободного от квадратов числа $d \in \mathbb{N}_1$ (другими словами, когда R изоморфно прямой сумме конечного числа простых полей различных характеристик).
4. Если группа $\mathcal{R}$ редуцирована и копериодична, то $R \in \mathfrak{D}$ тогда и только тогда, когда $R \cong \prod\limits_{p \in P} \mathbb{Z}_p$ для некоторого множества P простых чисел.

Доказательство. Докажем сначала импликации «только тогда». Предположим, что $R \in \mathfrak{D}$. Тогда, в частности, R — E-кольцо и $\Pi(R, 1) = \mathrm{RM}(\mathcal{R}, \mathcal{R})$ (см. утверждение 1). Если $\mathcal{R}$ периодична, то $R \cong \mathbb{Z}_d$ для некоторого $d \in \mathbb{N}_1$ согласно утверждению 3 леммы 5. Если $\mathcal{R}$ конечно порождена, то из замечания 8 следует, что $\mathcal{R}$ конечна, поэтому данный случай сводится к предыдущему. Если $\mathcal{R}$ не редуцирована, то $R \cong \mathbb{Z}_d \oplus \mathbb{Q}$ для некоторого $d \in \mathbb{N}_1$ ввиду утверждения 4 леммы 5. Пусть теперь $\mathcal{R}$ является прямой суммой подгрупп ранга 1. Тогда утверждение 5 леммы 5 показывает, что $R \cong \mathbb{Z}_d \oplus \bigoplus\limits_{i=1}^{k} \mathbb{Z}[1/X_i]$, где $d \in \mathbb{N}_1$, $k \in \mathbb{N}_0$, а $X_1, \ldots, X_k$ — множества простых чисел, содержащие все простые делители числа d, причём $X_i \nsubseteq X_j$ при $i \neq j$ $(i, j \in \{1, \ldots, k\})$. Из утверждения 2 леммы 2 вытекает, что $\Pi(\mathbb{Z}[1/X_i], 1) = \mathrm{RM}(\mathcal{Z}[1/X_i], \mathcal{Z}[1/X_i])$ для всех $i \in \{1, \ldots, k\}$. Лемма 7 показывает теперь, что $X_1 = \cdots = X_k = \mathbb{P}$. Следовательно, $k \leqslant 1$, так как $X_i \nsubseteq X_j$ при $i \neq j$. Таким образом, в рассматриваемом случае $R \cong \mathbb{Z}_d$ или $R \cong \mathbb{Z}_d \oplus \mathbb{Q}$ (напомним, что $\mathbb{Z}[1/\mathbb{P}] = \mathbb{Q}$).

Для завершения доказательства импликаций «только тогда» в утверждениях 1–3 настоящей теоремы осталось доказать, что во всех рассмотренных выше случаях чис-

ло d свободно от квадратов. Пусть $d = \prod\limits_{p\in D} p^{k_p}$, где D — конечное подмножество $\mathbb{P}$, а $k_p \in \mathbb{N}_1$ для всех $p \in D$. Тогда $\mathbb{Z}_d \cong \bigoplus\limits_{p\in D} \mathbb{Z}_{p^{k_p}}$. Пусть теперь $p \in D$. Снова применив утверждение 2 леммы 2, получаем, что $\Pi(\mathbb{Z}_{p^{k_p}}, 1) = \mathrm{RM}(\mathcal{Z}_{p^{k_p}}, \mathcal{Z}_{p^{k_p}})$. Из леммы 6 вытекает требуемое равенство $k_p = 1$.

Пусть теперь $\mathcal{R}$ редуцирована и копериодична. Тогда утверждение 6 леммы 5 показывает, что $R \cong \prod\limits_{p\in P} R_p$, где $P \subseteq \mathbb{P}$ и $R_p \in \{\mathbb{Z}_{p^k} : k \in \mathbb{N}_1\} \cup \{\mathbb{J}_p\}$ для каждого $p \in P$. Из утверждения 2 леммы 2 вытекает, что $\Pi(R_p, 1) = \mathrm{RM}(\mathcal{R}_p, \mathcal{R}_p)$ для всех $p \in P$. Следовательно, $R_p = \mathbb{Z}_p$ ввиду лемм 6 и 8. Таким образом, импликации «только тогда» доказаны.

Докажем теперь импликации «тогда». Очевидно, что $\mathrm{Hom}(\mathcal{Z}_p, \mathcal{Z}_q) = \{0\}$ для любых различных простых чисел p и q. Легко также видеть, что $\mathrm{Hom}(\mathcal{Q}, \mathcal{Z}_p) = \{0\}$ (так как группа $\mathcal{Q}$ делима) и $\mathrm{Hom}(\mathcal{Z}_p, \mathcal{Q}) = \{0\}$ (так как группа $\mathcal{Q}$ не имеет кручения) для каждого $p \in \mathbb{P}$. Поэтому из следствия 1 и леммы 3 вытекает, что $\bigoplus\limits_{p\in D} \mathbb{Z}_p \in \mathfrak{D}$ и $\left(\bigoplus\limits_{p\in D} \mathbb{Z}_p\right) \oplus \mathbb{Q} \in \mathfrak{D}$ для любого конечного множества D простых чисел. Отсюда следуют импликации «тогда» в утверждениях 1–3 теоремы.

Пусть теперь P — произвольное подмножество $\mathbb{P}$. Тогда для любого $p \in P$ группа $\prod\limits_{q\in P\setminus\{p\}} \mathcal{Z}_q$ p-делима и, следовательно, $\mathrm{Hom}\left(\prod_{q\in P\setminus\{p\}} \mathcal{Z}_q, \mathcal{Z}_p\right) = \{0\}$. Поэтому $\prod\limits_{p\in P} \mathbb{Z}_p \in \mathfrak{D}$ ввиду следствия 1 и леммы 3, что и требовалось. Импликации «тогда» доказаны. ■

Замечание 9. Доказательство теоремы 1 показывает, что указанные в её формулировке изоморфизмы имеют место, даже если R — E-кольцо и $\Pi(R, 1) = \mathrm{RM}(\mathcal{R}, \mathcal{R})$ (а не только если $R \in \mathfrak{D}$). Поэтому если группа $\mathcal{R}$ удовлетворяет хотя бы одному из условий, указанных в теореме 1, то $R \in \mathfrak{D}$ тогда и только тогда, когда R является E-кольцом и $\Pi(R, 1) = \mathrm{RM}(\mathcal{R}, \mathcal{R})$

ЛИТЕРАТУРА

1. *Логачев О. А., Сальников А. А., Ященко В. В.* Некоторые характеристики «нелинейности» групповых отображений // Дискретный анализ и исследование операций. Сер. 1. 2001. Т. 8. № 1. С. 40–54.
2. *Анохин М. И.* О некоторых множествах групповых функций // Матем. заметки. 2003. Т. 74. № 1. С. 3–11.
3. *Черемушкин А. В.* Аддитивный подход к определению степени нелинейности дискретной функции // Прикладная дискретная математика. 2010. № 2(8). С. 22–33.
4. *Черемушкин А. В.* Аддитивный подход к определению степени нелинейности дискретной функции на циклической группе примарного порядка // Прикладная дискретная математика. 2013. № 2(20). С. 26–38.
5. *Fuchs L.* Infinite Abelian Groups. Academic Press, 1970 (v. I), 1973 (v. II). Русс. пер.: *Фукс Л.* Бесконечные абелевы группы. М.: Мир, 1974 (т. 1), 1977 (т. 2).
6. *Fuchs L.* Abelian Groups. Springer, 2015.
7. *Schultz P.* The endomorphism ring of the additive group of a ring // J. Austral. Math. Soc. 1973. V. 15. No. 1. P. 60–69.
8. *Bowshell R. A. and Schultz P.* Unital rings whose additive endomorphisms commute // Math. Ann. 1977. V. 228. No. 3. P. 197–214.

9. *Riordan J.* Combinatorial identities. John Wiley & Sons, 1968. Русс. пер.: *Риордан Дж.* Комбинаторные тождества. М.: Наука, 1982.
10. *Jaballah A.* Subrings of **Q** // J. Sci. Technol. 1997. V. 2. No. 2. P. 1–13.
11. *Dimitrić R.* On coslender groups // Glasnik Matem. 1986. V. 21(41). No. 2. P. 327–329.

REFERENCES

1. *Logachev O. A., Sal'nikov A. A., and Yashchenko V. V.* Nekotorye kharakteristiki "nelineynosti" gruppovykh otobrazheniy [Some characteristics of "nonlinearity" of group mappings]. Diskretn. Anal. Issled. Oper., Ser. 1, 2001, vol. 8, no. 1, pp. 40–54. (in Russian)
2. *Anokhin M. I.* On some sets of group functions. Math. Notes, 2003, vol. 74, no. 1, pp. 3–11.
3. *Cheremushkin A. V.* Additivnyy podkhod k opredeleniyu stepeni nelineynosti diskretnoy funktsii [An additive approach to nonlinear degree of discrete function]. Prikladnaya Diskretnaya Matematika, 2010, no. 2(8), pp. 22–33. (in Russian)
4. *Cheremushkin A. V.* Additivnyy podkhod k opredeleniyu stepeni nelineynosti diskretnoy funktsii na tsiklicheskoy gruppe primarnogo poryadka [An additive approach to nonlinearity degree of discrete functions on a primary cyclic group]. Prikladnaya Diskretnaya Matematika, 2013, no. 2(20), pp. 26–38. (in Russian)
5. *Fuchs L.* Infinite Abelian Groups. Academic Press, 1970 (v. I), 1973 (v. II).
6. *Fuchs L.* Abelian Groups. Springer, 2015.
7. *Schultz P.* The endomorphism ring of the additive group of a ring. J. Austral. Math. Soc., 1973, vol. 15, no. 1, pp. 60–69.
8. *Bowshell R. A. and Schultz P.* Unital rings whose additive endomorphisms commute. Math. Ann., 1977, vol. 228, no. 3, pp. 197–214.
9. *Riordan J.* Combinatorial identities. John Wiley & Sons, 1968.
10. *Jaballah A.* Subrings of **Q**. J. Sci. Technol., 1997, vol. 2, no. 2, pp. 1–13.
11. *Dimitrić R.* On coslender groups. Glasnik Matem., 1986, vol. 21(41), no. 2, pp. 327–329.

UDC 512.772.7 DOI 10.17223/20710410/37/2

HYPERELLIPTIC CURVES, CARTIER — MANIN MATRICES AND LEGENDRE POLYNOMIALS

S. A. Novoselov

Immanuel Kant Baltic Federal University, Kaliningrad, Russia

Using hyperelliptic curves in cryptography requires the computation of the Jacobian order of a curve. This is equivalent to computing the characteristic polynomial of Frobenius $\chi(\lambda) \in \mathbb{Z}[\lambda]$. By calculating Cartier — Manin matrix, we can recover the polynomial $\chi(\lambda)$ modulo the characteristic of the base field. This information can further be used for recovering full polynomial in combination with other methods. In this paper, we investigate the hyperelliptic curves of the form $C_1 : y^2 = x^{2g+1} + ax^{g+1} + bx$ and $C_2 : y^2 = x^{2g+2} + ax^{g+1} + b$ over the finite field $\mathbb{F}_q$, $q = p^n$, $p > 2$. We transform these curves to the form $C_{1,\rho} : y^2 = x^{2g+1} - 2\rho x^{g+1} + x$ and $C_{2,\rho} : y^2 = x^{2g+2} - 2\rho x^{g+1} + 1$, where $\rho = -a/(2\sqrt{b})$, and prove that the coefficients of the corresponding Cartier — Manin matrices for the curves in this form are Legendre polynomials. As a consequence, the matrices are centrosymmetric and therefore, for finding the matrix, it's enough to compute a half of coefficients. Cartier — Manin matrices are determined up to a transformation of the form $S^{(p)}WS^{-1}$. It is known that centrosymmetric matrices can be transformed to the block-diagonal form by an orthogonal transformation. We prove that this transformation can be modified to have a form $S^{(p)}WS^{-1}$ and be defined over the base field of the curve. Therefore, Cartier — Manin matrices of curves $C_{1,\rho}$ and $C_{2,\rho}$ are equivalent to block-diagonal matrices. In the case of $\gcd(p, g) = 1$, Miller and Lubin proved that the matrices of curves C_1 and C_2 are monomial. We prove that the polynomial $\chi(\lambda) \pmod p$ can be found in factored form in terms of Legendre polynomials by using permutation attached to the monomial matrix. As an application of our results, we list all possible polynomials $\chi(\lambda) \pmod p$ in the case of $\gcd(p, g) = 1$, g is from 2 to 7 and the curve C_1 is over $\mathbb{F}_p$ if $\sqrt{b} \in \mathbb{F}_p$ and over $\mathbb{F}_{p^2}$ if $\sqrt{b} \notin \mathbb{F}_p$.

Keywords: *hyperelliptic curve cryptography, Cartier — Manin matrix, Legendre polynomials.*

Introduction

Let $\mathbb{F}_q$ be a finite field, $q = p^n$, $p > 2$. A hyperelliptic curve of a genus g over $\mathbb{F}_q$ is a nonsingular curve given by an equation

$$C : y^2 = f(x),$$

where $f \in \mathbb{F}_q[x]$, f is monic, $\deg f = 2g + 1$ or $\deg f = 2g + 2$.

Hyperelliptic curves were first proposed for use in cryptography by Koblitz [1]. Due to index-calculus attacks on hyperelliptic curves [2–4], only curves with a small genus are now considered in cryptography. In the more specific area of the cryptography on pairings, we are only interested in curves over prime and possibly medium or big characteristic fields, since in this case the security of cryptosystems relies on the discrete logarithm problem in finite fields, which has quasi-polynomial complexity for finite fields with a small characteristic [5].

The hyperelliptic curve C has an associated group — its Jacobian $J_C(\mathbb{F}_q)$, where all computations take place. For applications in cryptography, we need to compute the order

of $J_C(\mathbb{F}_q)$. Computing the order of Jacobian is equivalent to computing characteristic polynomial $\chi_q(\lambda)$ of the Frobenius endomorphism of J_C, which is determined by zeta function. If $N_k = \#C(\mathbb{F}_{q^k})$, then zeta function is a generating function

$$Z(\lambda) := \exp\left(\sum_{k=1}^{\infty} \frac{N_k}{k}\lambda^k\right) = \frac{L(\lambda)}{(1-\lambda)(1-q\lambda)},$$

where $L(\lambda) \in \mathbb{Z}[\lambda]$ is the L-polynomial, $L(\lambda) = \lambda^{2g}\chi_q(1/\lambda)$, and we have $\#J_C(\mathbb{F}_q) = \chi_q(1) = L(1)$.

Let $f(x)^{(p-1)/2} = \sum\limits_{i=0}^{(\deg f)(p-1)/2} c_i x^i$. Then Cartier — Manin matrix of the hyperelliptic curve C is a matrix

$$W = (w_{ij}) = \begin{pmatrix} c_{p-1} & c_{p-2} & \dots & c_{p-g} \\ c_{2p-1} & c_{2p-2} & \dots & c_{2p-g} \\ \dots & \dots & \dots & \dots \\ c_{gp-1} & c_{gp-2} & \dots & c_{gp-g} \end{pmatrix}.$$

Manin [6] showed that the characteristic polynomial of the matrix W is connected with the polynomial $\chi_q(\lambda)$ in the following way. Let $W_p = W \cdot W^{(p)} \cdot \ldots \cdot W^{(p^{n-1})}$, where $W^{(p^k)} = (w_{i,j}^{p^k})$, then

$$\chi_q(\lambda) \equiv (-1)^g \lambda^g |W_p - \lambda I_g| \pmod{p}.$$

Cartier — Manin matrices can in general be computed by optimized algorithms from [7, 8], which are faster than collecting coefficients after expansion of $f(x)^{(p-1)/2}$. After computing the polynomial $\chi_q(\lambda) \bmod p$, we can use Hasse — Weil bound in combination with other methods to recover full polynomial $\chi_q(\lambda)$.

In this work, we study hyperelliptic curves of the form

$$C_1 : y^2 = x^{2g+1} + ax^{g+1} + bx$$

and

$$C_2 : y^2 = x^{2g+2} + ax^{g+1} + b.$$

These curves are isomorphic to curves

$$C_{1,\rho} : y^2 = x^{2g+1} - 2\rho x^{g+1} + x$$

and

$$C_{2,\rho} : y^2 = x^{2g+2} - 2\rho x^{g+1} + 1$$

over the field $K = \mathbb{F}_q[\sqrt{b}]$. Therefore, we can restrict the discussion to the curves $C_{1,\rho}$ and $C_{2,\rho}$ and our results for polynomials $\chi(\lambda)$ hold over $\mathbb{F}_q$ if b is a square and over $\mathbb{F}_{q^2}$ if b is not a square in $\mathbb{F}_q$. These forms of curves are motivated by Jacobi quartics investigated by N. Yui [9].

The curves C_1 and C_2 were first studied by Miller and Lubin [10, 11], who proved that the Cartier — Manin matrices of these curves are the generalized permutation (monomial) matrices.

F. Leprevost and F. Morain [12] expressed the number of points of these curves in terms of certain modular functions, which can be efficiently computed for some special instances of curves.

For $g = 1$ these curves are elliptic ones. It is known that the number of points of elliptic curves in Legendre form for C_1 and Jacobi form for C_2 is congruent to Legendre polynomials

(see [9, 13] for details). Here, we show that this can be generalized to $g > 1$ case and prove that the number of points in J_{C_1} and J_{C_2} is congruent to an expression in terms of Legendre polynomials.

The genus 2 case was investigated for use in cryptography in [14 – 16]. It was proved that the genus 2 curves of the forms C_1 and C_2 have Jacobian isogenous to direct product of elliptic curves. Some explicit formulas for zeta function and for $\chi_q(\lambda)$ were found.

In this paper, we list all the possibilities for the polynomial $\chi_q(\lambda)$ modulo prime p for genus g from 2 to 7, $p > 2$, $\gcd(p, g) = 1$, and the curve C_1 over $\mathbb{F}_p$ if $b \in \mathbb{F}_p$ (Table 1) and over $\mathbb{F}_{p^2}$ if $b \notin \mathbb{F}_p$ (Table 2). Our methods can also be applied to any genus and finite field $\mathbb{F}_{p^n}$ with $\gcd(p, g) = 1$ and $p > 2$.

The rest of the paper is organized as follows. In section 1.1, we collect and prove preliminary results for monomial matrices and their permutations. In section 1.2, we prove necessary conditions for coefficients of Cartier — Manin matrices of C_1 and C_2 to be non-zero. From this, we also obtain conditions for the matrix to be diagonal or anti-diagonal.

In section 2.1, we prove that non-zero elements of Cartier — Manin matrix of the curve C_1 are Legendre polynomials and, as consequence, that the matrix is centrosymmetric. Using this fact, we prove that Cartier — Manin matrix of the curve C_1 is equivalent to a block-diagonal matrix over the finite field $\mathbb{F}_q$. In the case when the matrix is monomial with an attached permutation σ, we show how the polynomial $\chi_q(\lambda) \pmod p$ can be found in factored form by using this permutation and methods from Section 1.1. Section 2.2 contains analogous results for the curve C_2.

Tables 1 and 2 contain all the possible variants of the polynomials $\chi(\lambda) \pmod p$ for the case of $\gcd(g, p) = 1$, $p > 2$, and the curve C_1 over the fields $\mathbb{F}_p$ and $\mathbb{F}_{p^2}$.

1. Preliminary results

1.1. Permutations specified by congruence

A matrix M of size $n \times n$ is a generalized permutation (or monomial) matrix if each its column as well as each its row contains exactly one non-vanishing element. Every such a matrix can be decomposed into the product of a diagonal matrix and a permutation matrix

$$M = \operatorname{diag}(m_1, m_2, \ldots, m_n)P_\sigma$$

for some permutation $\sigma \in \operatorname{Sym}(n)$. Consider the case when the permutation σ is defined by a congruence modulo n.

Theorem 1. Let a, b, n be integers, $n > 1$, $a \not\equiv 1 \pmod n$, $\gcd(a, n) = 1$, $M = \operatorname{diag}(m_1, m_2, \ldots, m_n)P_\sigma$ be a monomial matrix, and σ be a permutation such that $\sigma(i) \equiv ai - b \pmod n$. Then

1) $\sigma^s(i) \equiv a^s i - b\left(\dfrac{a^s - 1}{a - 1}\right) \pmod n$;
2) $\operatorname{ord}(\sigma) = \operatorname{ord}_n(a)$;
3) if $d_j = \gcd(a^j - 1, n)$ and $b_j = b\left(\dfrac{a^j - 1}{a - 1}\right)$, then the number of cycles in the decomposition of the permutation σ into disjunct cycles equals

$$m = \frac{1}{\operatorname{ord}_n(a)}\left(n + \sum_{d_j | b_j} d_j\right),\ 1 \leqslant j \leqslant \operatorname{ord}_n(a) - 1;$$

4) if $\sigma = \sigma_1\sigma_2\dots\sigma_m$ is the disjunct cycles decomposition of σ, then the characteristic polynomial $\chi_M(\lambda)$ of the matrix M factors in the following way:

$$\chi_M(\lambda) = \prod_{j=1}^{m} (\lambda^{|\sigma_j|} - m_{\sigma_j}),$$

where m_{σ_j} is the product of all elements in the matrix M with indexes in the cycle σ_j.

Proof.

1) Let $s = 1$. Then $\sigma(i) \equiv ai - b\left(\dfrac{a-1}{a-1}\right) \pmod n$. Let $s+1 > 1$. Then

$$\sigma^{s+1}(i) = \sigma(\sigma^s(i)) \equiv a\left(a^s i - b\left(\frac{a^s-1}{a-1}\right)\right) - b \equiv a^{s+1}i - b\left(\frac{a^{s+1}-1}{a-1}\right) \pmod n.$$

So the formula is true by induction.

2) Let $r = \mathrm{ord}_n(a)$. Assume that there exists $j < r$ such that $\sigma^j(i) = i$ for all i. Then for all i, we have

$$(a^j - 1)i \equiv b\left(\frac{a^j-1}{a-1}\right) \pmod n.$$

This congruence has solutions iff $\gcd(a^j - 1, n) = d_j | b\left(\dfrac{a^j-1}{a-1}\right)$; in this case, the number of solutions is equal to d_j.

Since r is the minimal integer such that $a^r \equiv 1 \pmod n$, we have $a^j \not\equiv 1 \pmod n$. Then $d_j < n$ and there exists integer j_0 such that $\sigma^j(j_0) \not\equiv j_0 \pmod n$. This contradiction proves our statement.

3) Cycles in the disjunct decomposition of the permutation σ correspond to orbits in the action of the group $\langle\sigma\rangle$ on the set $S = \{1, \dots, n\}$.

The number of orbits can be calculated by Burnside's lemma:

$$m = \frac{1}{|\langle\sigma\rangle|}\sum_{j=1}^{|\langle\sigma\rangle|} \#\{i \in S : \sigma^j(i) = i\} = \frac{1}{r}\left(n + \sum_{j=1}^{r-1} \#\{i \in S : \sigma^j(i) = i\}\right).$$

The number of elements i such that $\sigma^j(i) = i$ is equal to the number of solutions of the congruence $a^j i - b\left(\dfrac{a^j-1}{a-1}\right) \equiv i \pmod n$, which is $d_j = \gcd(a^j - 1, n)$ if $d_j \Big| b\left(\dfrac{a^j-1}{a-1}\right)$ and 0 otherwise. Therefore,

$$m = \frac{1}{r}\left(n + \sum_{d_j | b_j} d_j\right).$$

4) See [17, Theorem 3]. ■

1.2. Hyperelliptic curves of the form $y^2 = x^t + ax^s + bx^m$

The next lemma gives some necessary conditions for coefficients of the Cartier — Manin matrix of a named form curve to be zero.

Lemma 1. Let $C : y^2 = x^t + ax^s + bx^m$ be a genus g hyperelliptic curve over finite field $\mathbb{F}_q$, $q = p^n$, $p > 2$, $t \in \{2g+2, 2g+1\}$, $m < s < t$, $m \in \{0, 1\}$ and $d = \gcd(t-m, s-m)$. Let $W = (w_{i,j})$, $1 \leqslant i, j \leqslant g$, be the Cartier — Manin matrix of the curve C. Then $w_{i,j} = 0$ for all i, j such that $ip - j \not\equiv m(p-1)/2 \pmod d$.

Proof. We have

$$w_{i,j} = [x^{ip-j}](x^t + ax^s + bx^m)^{(p-1)/2} = [x^{ip-j-m(p-1)/2}](x^{t-m} + ax^{s-m} + b)^{(p-1)/2} =$$
$$= [x^{ip-j-m(p-1)/2}] \sum_{k_1+k_2+k_3=(p-1)/2} \binom{(p-1)/2}{k_1, k_2, k_3} a^{k_2} b^{k_3} x^{(t-m)k_1+(s-m)k_2} = \sum_{k_1,k_2,k_3} \binom{(p-1)/2}{k_1, k_2, k_3} a^{k_2} b^{k_3},$$

where sum goes all k_1, k_2, k_3, which satisfy the system of equations

$$\begin{cases} k_1 + k_2 + k_3 = (p-1)/2, \\ (t-m)k_1 + (s-m)k_2 = ip - j - m(p-1)/2. \end{cases}$$

The second equation has a solution in integers k_1, k_2 if and only if $\gcd(t-m, s-m)$ divides $ip - j - m(p-1)/2$. Otherwise, the system has no solutions and we get $w_{i,j} = 0$. ■

From this lemma, we obtain some sufficient conditions for the Cartier — Manin matrix to be diagonal or anti-diagonal.

Theorem 2. Let $C : y^2 = x^{2g+1} + ax^{g+1} + bx$ be a genus g hyperelliptic curve over the finite field $\mathbb{F}_q$ and W be the Cartier — Manin matrix of this curve. Then

1) W is a diagonal matrix if one of the following conditions holds:
   a) g is even and $p \equiv 1 \pmod{2g}$;
   b) g is odd and $p \equiv 1 \pmod{g}$.
2) W is a anti-diagonal matrix if one of the following conditions holds:
   a) g is even and $p \equiv -1 \pmod{2g}$;
   b) g is odd and $p \equiv -1 \pmod{g}$.

Proof.

1) Let g be even and $p \equiv 1 \pmod{2g}$. Then $p = 1+2gm$ for some integer m. By Lemma 1 elements of matrix W can be non-zero only if $g|(ip - j - (p-1)/2) = i(1+2gm) - j - gm$, i.e. should be $i \equiv j \pmod{g}$. Since $1 \leqslant i, j \leqslant g$, we get $i = j$.

Let g be odd and $p \equiv 1 \pmod{g}$. Since $\gcd(g, 2) = 1$, $ip - j - (p-1)/2 \equiv i - j \pmod{g}$ and $i \equiv j \pmod{g}$.

2) The proof is similar to 1. ■

2. Main results

2.1. Curves of the form $y^2 = x^{2g+1} + ax^{g+1} + bx$

The genus g hyperelliptic curves of the form $C_1 : y^2 = x^{2g+1} + ax^{g+1} + bx$ over the finite field $\mathbb{F}_q$ are isomorphic over $\mathbb{F}_q[\sqrt{b}]$ to

$$C_{1,\rho} : y^2 = x^{2g+1} - 2\rho x^{g+1} + x, \ \rho = -\frac{a}{2\sqrt{b}}$$

via isomorphism

$$(x, y) \mapsto \left(b^{1/(2g)} x, b^{(2g+1)/(4g)} y\right).$$

Let $K = \mathbb{F}_q[\sqrt{b}]$. If b is a square in $\mathbb{F}_q$, then $K = \mathbb{F}_q$, otherwise $K \cong \mathbb{F}_{q^2}$.

First, we proof that the coefficients of the Cartier — Manin matrix W of the curve $C_{1,\rho}$ correspond to the Legendre polynomials.

Theorem 3. Let $C_{1,\rho} : y^2 = x^{2g+1} - 2\rho x^{g+1} + x$ be a genus g hyperelliptic curve over the finite field $\mathbb{F}_q$ and $W = (w_{i,j})$ be the Cartier — Manin matrix of $C_{1,\rho}$. Then

1) $w_{i,j} = 0$, if $ip - j \not\equiv (p-1)/2 \pmod{g}$;

2) $w_{i,j} \equiv P_{(ip-j)/g-(p-1)/(2g)}(\rho) \pmod p$, otherwise.

Proof.

1) The statement follows from Lemma 1.

2) Let $ip - j \equiv (p-1)/2 \pmod g$ and therefore $g|(ip - j - (p-1)/2)$. We have

$$w_{i,j} = \left[x^{ip-j-(p-1)/2}\right] (x^{2g} - 2\rho x^g + 1)^{(p-1)/2}.$$

Making substitution $z = x^g$, we get

$$w_{i,j} = \left[z^{(ip-j)/g-(p-1)/(2g)}\right] (z^2 - 2\rho z + 1)^{(p-1)/2} \equiv \left[z^{(ip-j)/g-(p-1)/(2g)}\right] \frac{1}{\sqrt{z^2 - 2\rho z + 1}} \pmod p.$$

Note that the generating function of the Legendre polynomials has the form

$$\sum_{k=0}^{\infty} P_k(x) z^k = \frac{1}{\sqrt{z^2 - 2xz + 1}}.$$

From this, it follows that $w_{i,j} \equiv P_{(ip-j)/g-(p-1)/(2g)}(\rho)$. ■

In many cases, the Cartier — Manin matrix of the curve $C_{1,\rho}$ has some special forms. We collect and prove these ones in the following theorem.

Theorem 4. Let $y^2 = x^{2g+1} - 2\rho x^{g+1} + x$ be a genus g hyperelliptic curve over the field $\mathbb{F}_q$ and W be the Cartier — Manin matrix of the curve. Then matrix W is

1) centrosymmetric in $\mathbb{F}_q$;
2) monomial, if $\gcd(p, g) = 1$;
3) diagonal, if one of the following conditions holds:
   a) g is even and $p \equiv 1 \pmod{2g}$;
   b) g is odd and $p \equiv 1 \pmod g$;
4) antidiagonal, if one of the following conditions holds:
   a) g is even and $p \equiv -1 \pmod{2g}$;
   a) g is odd and $p \equiv -1 \pmod g$.

Proof.

1) By Theorem 3, when $g|(ip - j - (p-1)/2)$ we have

$$w_{i,j} \equiv P_{(ip-j)/g-(p-1)/(2g)}(\rho) \pmod p.$$

From congruence properties of the Legendre polynomials [18, (5.9)], we get

$$P_{p-1-m}(\rho) \equiv P_m(\rho) \pmod p,\ 0 \leqslant m \leqslant p-1.$$

So

$$w_{i,j} \equiv P_{(ip-j)/g-(p-1)/(2g)}(\rho) \equiv P_{p-1-(ip-j)/g+(p-1)/(2g)}(\rho) \equiv$$
$$\equiv P_{((g-i+1)p-(g-j+1))/g-(p-1)/(2g)}(\rho) \equiv w_{g-i+1,g-j+1} \pmod p.$$

2) If j is fixed and $\gcd(p, g) = 1$, then the congruence $ip - j \equiv (p-1)/2 \pmod g$ has only one solution for i and since $1 \leqslant i \leqslant g$, there is only one j. Therefore, in every row, only one non-zero element is possible. Similarly, we can show that in every column, there can be only one non-zero element. From this, it follows that W is a monomial matrix.

3,4) See Theorem 2. ■

It's known that the set of centrosymmetric matrices and the set of monomial matrices are closed under multiplication of matrices. Note that if W is centrosymmetric (monomial), then $W^{(p^k)}$ is also centrosymmetric (monomial). Therefore matrix W_p is centrosymmetric (monomial), if W is centrosymmetric (monomial).

For centrosymmetric matrices, there is an orthogonal transformation [19], which transforms such matrices to block-diagonal form. If the size of a centrosymmetric matrix is even, than this transformation is defined by the non-singular orthogonal matrix

$$Q = \sqrt{\frac{1}{2}} \begin{pmatrix} I & -J \\ J & I \end{pmatrix}.$$

And for odd case

$$Q = \sqrt{\frac{1}{2}} \begin{pmatrix} I & 0 & -J \\ 0 & \sqrt{2} & 0 \\ J & 0 & I \end{pmatrix}.$$

Note that this transformation is defined over $\mathbb{F}_q[\sqrt{2}]$ and a different transformation is required for Cartier — Manin matrices. The Cartier — Manin matrix W of any hyperelliptic curve is determined up to transformation of the form $S^{(p)}WS^{-1}$, where S is a non-singular matrix [20, Proposition 2.2]. The following theorem shows that, by modifying transformation for centrosymmetric matrices, we can choose S in such way that the resulting matrix is block-diagonal and defined over $\mathbb{F}_q$.

Theorem 5. Let $C_{1,\rho}$ be a genus g hyperelliptic curve, defined by equation $y^2 = x^{2g+1} - 2\rho x^{g+1} + x$ over the finite field $\mathbb{F}_q$, $\operatorname{char}\mathbb{F}_q = p > 2$. Then the Cartier — Manin matrix W of $C_{1,\rho}$ is equivalent to a block-diagonal matrix.

Proof. Let $W = \begin{pmatrix} W_1 & W_3 \\ W_2 & W_4 \end{pmatrix}$ if g is even, and $W = \begin{pmatrix} W_1 & a & W_3 \\ b & c & d \\ W_2 & e & W_4 \end{pmatrix}$ if g is odd. Since, by Theorem 4, the matrix W is centrosymmetric in $\mathbb{F}_q$, then W can be written in the form $W = \begin{pmatrix} W_1 & JW_2J \\ W_2 & JW_1J \end{pmatrix}$ for even g and $W = \begin{pmatrix} W_1 & a & JW_2J \\ b & c & bJ \\ W_3 & Ja & JW_1J \end{pmatrix}$ if g is odd.

Consider the transformation of the form $S^{(p)}WS^{-1}$.

1. If $\sqrt{2} \in \mathbb{F}_q$, then we choose $S = Q$ and have $S^{(p)}WS^{-1} = Q^{(p)}WQ^{\mathrm{T}}$. We need to show that this transformation transforms matrix to the block-diagonal form.

If genus g is even, then $Q^{(p)} = \left(\frac{1}{2}\right)^{(p-1)/2} Q$ and

$$Q^{(p)}WQ^{\mathrm{T}} = \left(\frac{1}{2}\right)^{(p-1)/2} \begin{pmatrix} W_1 - JW_2 & 0 \\ 0 & J(W_1 + JW_2)J \end{pmatrix}.$$

If genus g is odd, then $Q^{(p)}WQ^{\mathrm{T}} = \left(\frac{1}{2}\right)^{(p-1)/2} \begin{pmatrix} W_1 - JW_2 & 0 & 0 \\ 0 & 2^{(p-1)/2}c & \sqrt{2}^p b \\ 0 & \sqrt{2}Ja & J(W_1 + JW_2)J \end{pmatrix}$.

2. Let $\sqrt{2} \notin \mathbb{F}_q$, choose $S = \begin{pmatrix} I & 0 & -J \\ 0 & 1 & 0 \\ J & 0 & I \end{pmatrix}$ for odd g and $S = \begin{pmatrix} I & -J \\ J & I \end{pmatrix}$ for even g.

Then we have $S^{(p)} = S$, since $p > 2$, and $S^{-1} = \frac{1}{2}\begin{pmatrix} I & 0 & J \\ 0 & 2 & 0 \\ -J & 0 & I \end{pmatrix}$ or $S^{-1} = \frac{1}{2}\begin{pmatrix} I & J \\ -J & I \end{pmatrix}$.

Now, we have

$$S^{(p)}WS^{-1} = \begin{pmatrix} W_1 - JW_2 & 0 & 0 \\ 0 & c & bJ \\ 0 & 2Ja & J(W_1 + JW_2)J \end{pmatrix}$$

for odd case and

$$S^{(p)}WS^{-1} = \begin{pmatrix} W_1 - JW_2 & 0 \\ 0 & J(W_1 + JW_2)J \end{pmatrix}$$

for even case.

Note that in this case the matrix S is not orthogonal. If the orthogonality of S is not required, this transformation can also be applied to the case $\sqrt{2} \in \mathbb{F}_q$. ■

Applying this transformation to W_p, we get a formula for the characteristic polynomial of the matrix W_p and therefore for $\chi_q(\lambda)$.

Corollary 1. Let $C_{1,\rho} : y^2 = x^{2g+1} - 2\rho x^{g+1} + x$ be a genus g hyperelliptic curve over the field $\mathbb{F}_q$, $q = p^n$, $p > 2$ and the matrix W be written in the above form. Then

1) if g is even,

$$\chi_q(\lambda) \equiv (-1)^g \lambda^g |(W_1 + JW_2)_p - \lambda I||(W_1 - JW_2)_p - \lambda I| \pmod{p};$$

2) if g is odd and $p|g$,

$$\chi_q(\lambda) \equiv (-1)^g \lambda^g \left| \begin{pmatrix} P_{(p-1)/2}(\rho) & bJ \\ 2Ja & J(W_1 + JW_2)J \end{pmatrix}_p - \lambda I \right| |(W_1 - JW_2)_p - \lambda I| \pmod{p};$$

3) if g is odd and $p \nmid g$,

$$\chi_q(\lambda) \equiv (-1)^g \lambda^g (N_{F_q/F_p}(P_{(p-1)/2}(\rho)) - \lambda)|(W_1 + JW_2)_p - \lambda I||(W_1 - JW_2)_p - \lambda I| \pmod{p}.$$

If the matrix W is monomial, we can go further.

Theorem 6. Let W be the Cartier — Manin matrix of the curve $C_{1,\rho}$ over the finite field $\mathbb{F}_q$, $\gcd(p, g) = 1$; σ be a permutation such that $\sigma(i) \equiv ip - (p-1)/2 \pmod{g}$, and $P(\sigma)$ be the permutation matrix for σ. Then

1) if g is even,

$$W = \operatorname{diag}(w_{1,\sigma(1)}, \ldots, w_{g/2,\sigma(g/2)}, w_{g/2+1,g/2+1-\sigma(g/2)}, w_{g/2+2,g/2+1-\sigma(g/2-1)}, \ldots, w_{g,g+1-\sigma(1)})P(\sigma);$$

2) if g is odd,

$$W = \operatorname{diag}(w_{1,\sigma(1)}, \ldots, w_{(g-1)/2,\sigma((g-1)/2)}, w_{(g+1)/2,(g+1)/2}, w_{(g+1)/2+1,g+1-\sigma((g+1)/2-1)}, \ldots, w_{g,g+1-\sigma(1)})P(\sigma).$$

Proof. If $\gcd(p, g) = 1$, then W is a monomial matrix, which can be factored in the product of diagonal and permutation matrix: $W = \operatorname{diag}(w_{1,\sigma(1)}, \ldots, w_{g,\sigma(g)})P(\sigma)$. By Lemma 1, for non-zero elements of W, we have $ip - j \equiv (p-1)/2 \pmod{g}$. So the permutation σ is defined as $\sigma(i) \equiv ip - (p-1)/2 \pmod{g}$. Since the matrix $P(\sigma)$ is also centrosymmetric, every i such that $\sigma(i) \equiv ip - (p-1)/2 \pmod{g}$ uniquely determines the value of $\sigma(g+1-i)$, as $\sigma(g+1-i) \equiv g+1-\sigma(i) \pmod{g}$. ■

If we know the decomposition of σ^n into disjoint cycles, we can factor the polynomial $\chi_q(\lambda)$ in the following way.

Theorem 7. Let $C_{1,\rho} : y^2 = x^{2g+1} - 2\rho x^{g+1} + x$ be a hyperelliptic curve over the finite field $\mathbb{F}_q$, $q = p^n$, $\gcd(p, g) = 1$ and W be the Cartier — Manin matrix of this curve. Then W is the monomial matrix with the permutation σ such that $\sigma(i) \equiv ip - (p-1)/2 \pmod g$ and W_p is a monomial matrix with the permutation σ^n such that $\sigma^n(i) \equiv ip^n(p^n-1)/2 \pmod g$. If $W_p = (w'_{i,j})$ and $\sigma^n = \sigma_1\sigma_2 \dots \sigma_m$ is the decomposition of σ^n into disjoint cycles, then

$$\chi_q(\lambda) \equiv \lambda^g \prod_{j=1}^{m} (\lambda^{|\sigma_j|} - \prod_{k=1}^{|\sigma_j|} w'_{\sigma_{j,k},\sigma_{j,k+1}}) \pmod p,$$

where $\sigma_{j,k} = j_k$ for $\sigma_j = (j_1, \dots, j_{|\sigma_j|})$.

Proof. If W is the monomial matrix with the permutation σ, then, by multiplying matrices, we obtain

$$W_p = (w'_{i,j}) = \left(w^{p^{n-1}}_{\sigma^{n-1}(i),j} \prod_{k=0}^{n-2} w^{p^k}_{\sigma^k(i),\sigma^{k+1}(i)} \right),$$

where σ^k are permutations with $\sigma^k(i) \equiv ip^k - (p^k-1)/2 \pmod g$ and $w'_{i,j} = 0$ for all $j \neq \neq \sigma^n(i)$. Therefore, W_p is a monomial matrix with the permutation $\sigma^n(i) \equiv ip^n - (p^n-1)/2 \pmod g$ and the result follows from the Theorem 1. ■

In the case of the diagonal matrix, the formula can be made simpler.

Theorem 8. Let $C_{1,\rho} : y^2 = x^{2g+1} - 2\rho x^{g+1} + x$ be a genus g hyperelliptic curve over the finite field $\mathbb{F}_q$, $q = p^n$, $p > 2$. Then

1) if g is even and $p \equiv 1 \pmod{2g}$,

$$\chi_q(\lambda) \equiv \lambda^g \prod_{i=1}^{g/2} (\lambda - N_{\mathbb{F}_q/\mathbb{F}_p}(P_{(2i-1)(p-1)/(2g)}(\rho)))^2 \pmod p;$$

2) if g is odd and $p \equiv 1 \pmod g$,

$$\chi_q(\lambda) \equiv \lambda^g (\lambda - N_{\mathbb{F}_q/\mathbb{F}_p}(P_{(p-1)/2}(\rho))) \prod_{i=1}^{(g-1)/2} (\lambda - N_{\mathbb{F}_q/\mathbb{F}_p}(P_{(2i-1)(p-1)/(2g)}(\rho)))^2 \pmod p.$$

2.2. Curves of the form $y^2 = x^{2g+2} + ax^{g+1} + b$

The following curves of this form

$$C_2 : y^2 = x^{2g+2} + ax^{g+1} + b,$$
$$C_{2,\rho} : y^2 = x^{2g+2} - 2\rho x^{g+1} + 1$$

have the properties similar to the curves C_1 and $C_{1,\rho}$. We collect them in the following theorem.

Theorem 9. Let $C_{2,\rho}$ be a hyperelliptic curve defined by the equation $y^2 = x^{2g+2} - \\ - 2\rho x^{g+1} + 1$ over finite field $\mathbb{F}_q$ and $W = (w_{i,j})$ be the Cartier — Manin matrix of $C_{2,\rho}$. Then

1) $w_{i,j} = 0$ if $ip \not\equiv j \pmod{g+1}$;
2) $w_{i,j} \equiv P_{(ip-j)/(g+1)}(\rho) \pmod{g+1}$;
3) W is a centrosymmetric matrix in $\mathbb{F}_q$;
4) W is a monomial matrix if $p \nmid (g+1)$;
5) W is a diagonal matrix if $p \nmid (g+1)$ and $p \equiv 1 \pmod{g+1}$;

6) W is an anti-diagonal matrix if $p \nmid (g+1)$ and $p \equiv -1 \pmod{g+1}$;
7) there is a transformation of the form $S^{(p)} W S^{-1}$ where S is non-singular, which transforms W to a block-diagonal form;
8) if g is even,

$$\chi_q(\lambda) \equiv (-1)^g \lambda^g |(W_1 + JW_2)_p - \lambda I||(W_1 - JW_2)_p - \lambda I| \pmod{p};$$

9) if g is odd and $p|g$,

$$\chi_q(\lambda) \equiv (-1)^g \lambda^g \left| \begin{pmatrix} P_{(p-1)/2}(\rho) & bJ \\ 2Ja & J(W_1 + JW_2)J \end{pmatrix}_p - \lambda I \right| |(W_1 - JW_2)_p - \lambda I| \pmod{p};$$

10) if g is odd and $p \nmid g$,

$$\chi_q(\lambda) \equiv (-1)^g \lambda^g (N_{F_q/F_p}(P_{(p-1)/2}(\rho)) - \lambda) |(W_1 + JW_2)_p - \lambda I||(W_1 - JW_2)_p - \lambda I| \pmod{p}.$$

Proof.
1) It follows from Lemma 1.
2) Let $(g+1)|(ip-j)$ and $t = x^{g+1}$. Then

$$w_{i,j} = [x^{ip-j}](x^{2g+2} - 2\rho x^{g+1} + 1)^{(p-1)/2} = [t^{(ip-j)/(g+1)}](t^2 - 2\rho t + 1)^{(p-1)/2} \equiv P_{(ip-j)/(g+1)}(\rho) \pmod{p}.$$

3) $w_{i,j} \equiv P_{(ip-j)/(g+1)} \equiv P_{p-1-(ip-j)/(g+1)} \equiv w_{g+1-i,g+1-j}$.
4) If $\gcd(p, g+1) = 1$, then the congruence $ip \equiv j \pmod{g+1}$ has only one solution for each i, j, and since $1 \leqslant i, j \leqslant g$ it uniquely determines i, j.
5, 6) These follow from congruences $i \equiv j \pmod{g+1}$ and $i \equiv -j \pmod{g+1}$ for $1 \leqslant i, j \leqslant g$.
7–10) The needed transformations are taken from the Theorem 5. ■

Conclusion

We have proved that the Cartier — Manin matrices W for the curves $C_{1,\rho}$ and $C_{2,\rho}$ have a very special form, namely, the coefficients of W are the Legendre polynomials, W is centrosymmetric and is equivalent to a block-diagonal matrix. In the case $\gcd(p, g) = 1$, the matrices of C_1 and C_2 are monomial. Using this fact, we have proved (Theorem 7) that the polynomial $\chi_q(\lambda)$ modulo p can be computed in a factored form in terms of the Legendre polynomials. The matrix symmetry can be used to speed up the algorithms for computing the Cartier — Manin matrices, because it is enough to compute half of coefficients to completely determine a matrix itself. As an application, we have listed all the possible variants of the polynomial $\chi_p(\lambda)$ modulo p for the curve C_1 over prime field (Table 1) and over $\mathbb{F}_{p^2}$ (Table 2).

T a b l e 1

Hyperelliptic curves of the form $C_{1,\rho} : y^2 = x^{2g+1} + ax^{g+1} + bx$ over the prime field $\mathbb{F}_p$, $p > 2$, $p \nmid g$, $P_m := P_m(\rho)$ and b is a square

g	Conditions	$\chi_p(\lambda) \pmod{p}$
2	$p \equiv 1 \pmod 4$	$\lambda^2(\lambda - P_{(p-1)/4})^2$
2	$p \equiv 3 \pmod 4$	$\lambda^2(\lambda^2 - P^2_{(p-3)/4})$
3	$p \equiv 1 \pmod 3$	$\lambda^3(\lambda - P_{(p-1)/2})(\lambda - P_{(p-1)/6})^2$
3	$p \equiv 2 \pmod 3$	$\lambda^3(\lambda - P_{(p-1)/2})(\lambda^2 - P^2_{(p-5)/6})$
4	$p \equiv 1 \pmod 8$	$\lambda^4(\lambda - P_{(p-1)/8})^2(\lambda - P_{(3p-3)/8})^2$
4	$p \equiv 3 \pmod 8$	$\lambda^4(\lambda^2 - P_{(p-3)/8}P_{(3p-1)/8})^2$

End of Table 1

g	Conditions	$\chi_p(\lambda) \pmod p$
4	$p \equiv 5 \pmod 8$	$\lambda^4(\lambda^2 - P_{(p-5)/8}P_{(3p-7)/8})^2$
4	$p \equiv 7 \pmod 8$	$\lambda^4(\lambda^2 - P^2_{(p-7)/8})(\lambda^2 - P^2_{(3p-5)/8})$
5	$p \equiv 1 \pmod 5$	$\lambda^5(\lambda - P_{(p-1)/2})(\lambda - P_{(p-1)/10})^2(\lambda - P_{(3p-3)/10})^2$
5	$p \equiv 2 \pmod 5$	$\lambda^5(\lambda - P_{(p-1)/2})(\lambda^4 - P^2_{(p-7)/10}P^2_{(3p-1)/10})$
5	$p \equiv 3 \pmod 5$	$\lambda^5(\lambda - P_{(p-1)/2})(\lambda^4 - P^2_{(p-3)/10}P^2_{(3p-9)/10})$
5	$p \equiv 4 \pmod 5$	$\lambda^5(\lambda - P_{(p-1)/2})(\lambda^2 - P^2_{(p-9)/10})(\lambda^2 - P^2_{(3p-7)/10})$
6	$p \equiv 1 \pmod{12}$	$\lambda^6(\lambda - P_{(p-1)/12})^2(\lambda - P_{(p-1)/4})^2(\lambda - P_{(5p-5)/12})^2$
6	$p \equiv 5 \pmod{12}$	$\lambda^6(\lambda - P_{(p-1)/4})^2(\lambda^2 - P_{(p-5)/12}P_{(5p-1)/12})^2$
6	$p \equiv 7 \pmod{12}$	$\lambda^6(\lambda^2 - P_{(p-7)/12}P_{(5p-11)/12})^2(\lambda^2 - P^2_{(p-3)/4})$
6	$p \equiv 11 \pmod{12}$	$\lambda^6(\lambda^2 - P^2_{(p-11)/12})(\lambda^2 - P^2_{(p-3)/4})(\lambda^2 - P^2_{(5p-7)/12})$
7	$p \equiv 1 \pmod 7$	$\lambda^7(\lambda - P_{(p-1)/2})(\lambda - P_{(p-1)/14})^2(\lambda - P_{(3p-3)/14})^2(\lambda - P_{(5p-5)/14})^2$
7	$p \equiv 2 \pmod 7$	$\lambda^7(\lambda - P_{(p-1)/2})(\lambda^3 - P_{(p-9)/14}P_{(3p-13)/14}P_{(5p-3)/14})^2$
7	$p \equiv 3 \pmod 7$	$\lambda^7(\lambda - P_{(p-1)/2})(\lambda^6 - P^2_{(p-3)/14}P^2_{(3p-9)/14}P^2_{(5p-1)/14})$
7	$p \equiv 4 \pmod 7$	$\lambda^7(\lambda - P_{(p-1)/2})(\lambda^3 - P_{(p-11)/14}P_{(3p-5)/14}P_{(5p-13)/14})^2$
7	$p \equiv 5 \pmod 7$	$\lambda^7(\lambda - P_{(p-1)/2})(\lambda^6 - P^2_{(5p-11)/14}P^2_{(3p-1)/14}P^2_{(p-5)/14})$
7	$p \equiv 6 \pmod 7$	$\lambda^7(\lambda - P_{(p-1)/2})(\lambda^2 - P^2_{(p-13)/14})(\lambda^2 - P^2_{(3p-11)/14})(\lambda^2 - P^2_{(5p-9)/14})$

Table 2

Hyperelliptic curves of the form $C_{1,\rho} : y^2 = x^{2g+1} + ax^{g+1} + bx$ over the field $\mathbb{F}_{p^2}$, $p > 2$, $p \nmid g$, $P_m := P_m(\rho)$, b is a square in $\mathbb{F}_{p^2}$

g	Conditions	$\chi_{p^2}(\lambda) \pmod p$
2	$p \equiv 1 \pmod 4$	$\lambda^2(\lambda - P^{p+1}_{(p-1)/4})^2$
2	$p \equiv 3 \pmod 4$	$\lambda^2(\lambda - P^{p+1}_{(p-3)/4})^2$
3	$p \equiv 1 \pmod 3$	$\lambda^3(\lambda - P^{p+1}_{(p-1)/2})(\lambda - P^{p+1}_{(p-1)/6})^2$
3	$p \equiv 2 \pmod 3$	$\lambda^3(\lambda - P^{p+1}_{(p-1)/2})(\lambda - P^{p+1}_{(p-5)/6})^2$
4	$p \equiv 1 \pmod 8$	$\lambda^4(\lambda - P^{p+1}_{(p-1)/8})^2(\lambda - P^{p+1}_{(3p-3)/8})^2$
4	$p \equiv 3 \pmod 8$	$\lambda^4(\lambda - P^{p}_{(p-3)/8}P_{(3p-1)/8})^2(\lambda - P^{p}_{(3p-1)/8}P_{(p-3)/8})^2$
4	$p \equiv 5 \pmod 8$	$\lambda^4(\lambda - P^{p}_{(p-5)/8}P_{(3p-7)/8})^2(\lambda - P^{p}_{(3p-7)/8}P_{(p-5)/8})^2$
4	$p \equiv 7 \pmod 8$	$\lambda^4(\lambda - P^{p+1}_{(p-7)/8})^2(\lambda - P^{p+1}_{(3p-5)/8})^2$
5	$p \equiv 1 \pmod 5$	$\lambda^5(\lambda - P^{p+1}_{(p-1)/2})(\lambda - P^{p+1}_{(p-1)/10})^2(\lambda - P^{p+1}_{(3p-3)/10})^2$
5	$p \equiv 2 \pmod 5$	$\lambda^5(\lambda^2 - P^{2p}_{(p-7)/10}P^2_{(3p-1)/10})(\lambda^2 - P^{2p}_{(3p-1)/10}P^2_{(p-7)/10})(\lambda - P^{p+1}_{(p-1)/2})$
5	$p \equiv 3 \pmod 5$	$\lambda^5(\lambda^2 - P^{2p}_{(p-3)/10}P^2_{(3p-9)/10})(\lambda^2 - P^{2p}_{(3p-9)/10}P^2_{(p-3)/10})(\lambda - P^{p+1}_{(p-1)/2})$
5	$p \equiv 4 \pmod 5$	$\lambda^5(\lambda - P^{p+1}_{(p-1)/2})(\lambda - P^{p+1}_{(p-9)/10})^2(\lambda - P^{p+1}_{(3p-7)/10})^2$
6	$p \equiv 1 \pmod{12}$	$\lambda^6(\lambda - P^{p+1}_{(p-1)/4})^2(\lambda - P^{p+1}_{(p-1)/12})^2(\lambda - P^{p+1}_{(5p-5)/12})^2$
6	$p \equiv 5 \pmod{12}$	$\lambda^6(\lambda - P^{p}_{(p-5)/12}P_{(5p-1)/12})^2(\lambda - P^{p}_{(5p-1)/12}P_{(p-5)/12})^2(\lambda - P^{p+1}_{(p-1)/4})^2$
6	$p \equiv 7 \pmod{12}$	$\lambda^6(\lambda - P^{p}_{(p-7)/12}P_{(5p-11)/12})^2(\lambda - P^{p}_{(5p-11)/12}P_{(p-7)/12})^2(\lambda - P^{p+1}_{(p-3)/4})^2$
6	$p \equiv 11 \pmod{12}$	$\lambda^6(\lambda - P^{p+1}_{(p-3)/4})^2(\lambda - P^{p+1}_{(p-11)/12})^2(\lambda - P^{p+1}_{(5p-7)/12})^2$
7	$p \equiv 1 \pmod 7$	$\lambda^7(\lambda - P^{p+1}_{(p-1)/2})(\lambda - P^{p+1}_{(p-1)/14})^2(\lambda - P^{p+1}_{(5p-5)/14})^2(\lambda - P^{p+1}_{(3p-3)/14})^2$
7	$p \equiv 2 \pmod 7$	$\lambda^7(\lambda - P^{p+1}_{(p-1)/2})(\lambda^3 - P^{p+1}_{(p-9)/14}P^{p+1}_{(5p-3)/14}P^{p+1}_{(3p-13)/14})^2$
7	$p \equiv 3 \pmod 7$	$\lambda^7(\lambda - P^{p+1}_{(p-1)/2})(\lambda^3 - P^{p+1}_{(p-3)/14}P^{p+1}_{(3p-9)/14}P^{p+1}_{(5p-1)/14})^2$
7	$p \equiv 4 \pmod 7$	$\lambda^7(\lambda - P^{p+1}_{(p-1)/2})(\lambda^3 - P^{p+1}_{(p-11)/14}P^{p+1}_{(3p-5)/14}P^{p+1}_{(5p-13)/14})^2$
7	$p \equiv 5 \pmod 7$	$\lambda^7(\lambda^3 - P^{p+1}_{(p-5)/14}P^{p+1}_{(3p-1)/14}P^{p+1}_{(5p-11)/14})^2(\lambda - P^{p+1}_{(p-1)/2})$
7	$p \equiv 6 \pmod 7$	$\lambda^7(\lambda - P^{p+1}_{(p-1)/2})(\lambda - P^{p+1}_{(p-13)/14})^2(\lambda - P^{p+1}_{(3p-11)/14})^2(\lambda - P^{p+1}_{(5p-9)/14})^2$

Our results were checked in Pari/GP and Sage.

A short information about these results were presented by the author on the conference Sibecrypt'17 [21].

REFERENCES

1. *Koblitz N.* Hyperelliptic cryptosystems. J. Cryptology, 1989, vol. 1, no. 3, pp. 139–150.
2. *Enge A. and Gaudry P.* A general framework for subexponential discrete logarithm algorithms. Acta Arith., 2000, vol. 102, pp. 83–103.
3. *Enge A., Gaudry P., and Thomé E.* An $L(1/3)$ discrete logarithm algorithm for low degree curves. J. Cryptology, 2011, vol. 24, no. 1, pp. 24–41.
4. *Gaudry P., Thomé E., Thériault N., and Diem C.* A double large prime variation for small genus hyperelliptic index calculus. Math. Comput., 2007, vol. 76, no. 257, pp. 475–492.
5. *Barbulescu R., Gaudry P., Joux A., and Thomé E.* A heuristic quasi-polynomial algorithm for discrete logarithm in finite fields of small characteristic. LNCS, 2014, vol. 8441, pp. 1–16.
6. *Manin Yu. I.* O matritse Khasse — Vitta algebraicheskoy krivoy [The Hasse-Witt matrix of an algebraic curve]. Izv. Akad. Nauk SSSR, Ser. Mat., 1961, vol. 25, no. 1, pp. 153–172. (in Russian)
7. *Bostan A., Gaudry P., and Schost; E.* Linear recurrences with polynomial coefficients and application to integer factorization and Cartier — Manin operator. SIAM J. Comput., 2007, vol. 36, no. 6, pp. 1777–1806.
8. *Harvey D. and Sutherland A. V.* Hasse — Witt matrices of hyperelliptic curves in average polynomial time. LMS J. Comput. Math., 2014, vol. 17, no. A, pp. 257–273.
9. *Yui N.* Jacobi quartics, Legendre polynomials and formal groups. Lecture Notes in Mathematics, 1988, vol. 1326, pp. 182–215.
10. *Miller L.* The Hasse — Witt-matrix of special projective varieties. Pacific J. Math., 1972, vol. 43, no. 2, pp. 443–455.
11. *Miller L.* Curves with invertible Hasse — Witt-matrix. Math. Ann., 1972, vol. 197, pp. 123–127.
12. *Leprevost F. and Morain F.* Revêtements de courbes elliptiques à multiplication complexe par des courbes hyperelliptiques et sommes de caractères. J. Number Theory, 1997, vol. 64, no. 2, pp. 165–182.
13. *Brillhart J. and Morton P.* Class numbers of quadratic fields, Hasse invariants of elliptic curves, and the supersingular polynomial. J. Number Theory, 2004, vol. 106, no. 1, pp. 79–111.
14. *Satoh T.* Generating genus two hyperelliptic curves over large characteristic finite fields. LNCS, 2009, vol. 5479, pp. 536–553.
15. *Freeman D. M. and Satoh T.* Constructing pairing-friendly hyperelliptic curves using Weil restriction. J. Number Theory, 2011, vol. 131, no. 5, pp. 959–983.
16. *Guillevic A. and Vergnaud D.* Genus 2 hyperelliptic curve families with explicit Jacobian order evaluation and pairing-friendly constructions. LNCS, 2012, vol. 7708, pp. 234–253.
17. *Garcia-Planas M. I. and Magret M. D.* Eigenvectors of permutation matrices. Adv. Pure Math., 2015, vol. 5, no. 7, pp. 390–393.
18. *Carlitz L.* Congruence properties of the polynomials of Hermite, Laguerre and Legendre. Mathematische Zeitschrift, 1953, vol. 59, pp. 474–483.
19. *Weaver J. R.* Centrosymmetric (cross-symmetric) matrices, their basic properties, eigenvalues, and eigenvectors. Amer. Math. Monthly, 1985, vol. 92, no. 10, pp. 711–717.
20. *Yui N.* On the Jacobian varieties of hyperelliptic curves over fields of characteristic $p > 2$. J. Algebra, 1978, vol. 52, no. 2, pp. 378–410.
21. *Novoselov S. A.* Hyperelliptic curves, Cartier — Manin matrices and Legendre polynomials. Prikladnaya Diskretnaya Matematika. Prilozhenie, 2017, no. 10, pp. 30–32.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ КРИПТОГРАФИИ

УДК 519.17

О ПРИМИТИВНОСТИ ПЕРЕМЕШИВАЮЩИХ ОРГРАФОВ РЕГИСТРОВ СДВИГА С ДВУМЯ ОБРАТНЫМИ СВЯЗЯМИ

А. М. Коренева

Национальный исследовательский ядерный университет «МИФИ», г. Москва, Россия

С помощью матрично-графового подхода исследуются перемешивающие свойства преобразований регистров сдвига с двумя обратными связями над множеством V_r двоичных r-мерных векторов, $r > 1$. Под перемешивающими свойствами понимается существенная зависимость координатных булевых функций различных степеней регистровых преобразований от знаков начального состояния регистра, рассматриваемых как независимые переменные. Для перемешивающих орграфов подстановок регистров сдвига с двумя обратными связями, построенных на основе модифицированных аддитивных генераторов, доказан критерий примитивности и получены достижимые верхние оценки экспонента, которые существенно улучшают все другие известные оценки экспонентов для тех же орграфов.

Ключевые слова: *матрично-графовый подход, модифицированный аддитивный генератор, перемешивающий орграф, примитивность, регистр сдвига, экспонент.*

DOI 10.17223/20710410/37/3

ON PRIMITIVITY OF MIXING DIGRAPHS ASSOCIATED WITH 2-FEEDBACKS SHIFT REGISTERS

A. M. Koreneva

National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russia

E-mail: alisa.koreneva@gmail.com

Analysis of mixing properties of round transformations is an important issue in the theory of symmetric iterative block ciphers. For researching this subject, a matrix-digraph approach is widely used in cryptography. This approach allows to characterize the required properties in terms of primitivity and exponent of a matrix (or a digraph) related to the transformations concerned. This paper is devoted to such a characterization of mixing properties of transformations fulfilled by 2-feedback shift registers. For naturals n, m, and r, let $n > 1$, $r > 1$, $0 \leqslant m \leqslant n-2$, $V_r = (\mathrm{GF}(2))^r$; $f_m : V_r^n \to V_r$ and $f_{n-1} : V_r^n \to V_r$ are some feedback functions; $\mu : V_r \to V_r$ and $g : V_r \to V_r$ are some permutations over V_r used to modify feedbacks f_m and f_{n-1} respectively; $x_{\delta_0}, x_{\delta_1}, \ldots, x_{\delta_p}$ are all essential variables of the function $f_m(x_0, x_1, \ldots, x_{n-1})$, $\delta_0 = m+1$, $0 < \delta_1 < \ldots < \delta_p < n$, $p > 0$; $x_{d_0}, x_{d_1}, \ldots, x_{dq}$ are all essential variables of the function $f_{n-1}(x_0, x_1, \ldots, x_{n-1})$,

$d_0 = 0$, $d_1 < \ldots < d_q < n$, $q > 0$; $\varphi^{g,\mu} : V_r^n \to V_r^n$, $\varphi^{g,\mu}(x_0, x_1, \ldots, x_{n-1}) = (x_1, \ldots, x_{m-1}, \mu(f_m(x_0, \ldots, x_{n-1})), x_{m+1}, \ldots, x_{n-2}, g(f_{n-1}(x_0, \ldots, x_{n-1})))$. In fact, $\varphi^{g,\mu}$ is the transition function of a shift register of the length n over V_r with two feedback functions $\mu(f_m(x))$ and $g(f_{n-1}(x))$, $x = x_0x_1 \ldots x_{n-1}$. Let $M(\varphi^{g,\mu}) = M$ be a Boolean matrix (m_{ij}) (called the mixing matrix of the map $\varphi^{g,\mu}$), where $m_{ij} = 1$ iff the j-th coordinate function of the map $\varphi^{g,\mu}$ essentially depends on the variable x_i $(i, j \in \{0, 1, \ldots, n-1\})$. The matrix M is said to be primitive if there is a power $M^e = \left(m_{ij}^{(e)}\right)$ of its mixing matrix M such that $m_{ij}^{(e)} > 0$ for all i and j; in this case, the least power e is called an exponent of M and is denoted by $\exp M$. The conceptions of the primitiveness and exponent of the matrix $M(\varphi^{g,\mu})$ expend to the digraph $\Gamma(\varphi^{g,\mu})$ with the adjacency matrix M — the mixing graph associated with $\varphi^{g,\mu}$. The main results of the paper are the following: 1) it is proved that the strongly connected digraph $\Gamma(\varphi^{g,\mu})$ is primitive iff $\delta_1 > m$ and the numbers in the set $L' = \{n - d_i, n + m + 1 - d_j - \delta_k : i = 0, \ldots, q, j = 0, \ldots, t, k = 1, \ldots, p\}$ are relatively prime or $\delta_1 \leqslant m$ and the numbers in the set $L = \{n - d_i, n + m + 1 - d_j - \delta_k, m + 1 - \delta_l : i = 0, \ldots, q, j = 0, \ldots, t, k = \tau + 1, \ldots, p, l = 1, \ldots, \tau\}$ are relatively prime, where t and τ are determined by the conditions: d_t and δ_τ are the largest numbers in $D = \{d_0, \ldots, d_q\}$ and $\Delta = \{\delta_0, \ldots, \delta_p\}$ with the properties $d_t \leqslant m$ and $\delta_\tau \leqslant m$ respectively; 2) for $\exp \Gamma(\varphi^{g,\mu})$, some attainable upper bounds depending on m and other parameters in D and Δ are obtained, improving all the known exponent estimates for the same digraphs. Particularly, if $(n-1) \in D$ and $m \in \Delta$, then $\exp \Gamma(\varphi^{g,\mu}) \leqslant \min\{\rho(D) + \varepsilon, \rho(\Delta) + \varepsilon'\}$, where $\rho(D) = \max\{n - d_q, d_q - d_{q-1}, \ldots, d_1 - d_0\}$, $\rho(\Delta) = \max\{\delta_1 + n - \delta_p, \delta_p - \delta_{p-1}, \ldots, \delta_0 - \delta_r, \ldots, \delta_2 - \delta_1\}$, $\varepsilon = \max\{2n - m - 2 - d_q, n + m - \max\{\delta_0, \delta_p\}\}$, and $\varepsilon' = \max\{2m + 1 - \delta_\tau, n - 1 - d_t\}$. These results can be successfully used in construction of iterative cryptographic algorithms based on $\varphi^{g,\mu}$ with the rapid input data mixing.

Keywords: *primitive digraph, exponent, mixing digraph, multi-feedback shift register, modified additive generator.*

Введение

Введём основные обозначения:

- V_n — n-мерное пространство двоичных векторов, $n \in \mathbb{N}$, $n > 1$;
- $\mathbb{Z}_n$ — кольцо вычетов по модулю n, $n > 1$;
- $E(\varphi)$ — множество номеров существенных переменных дискретной функции φ;
- $\exp \Gamma$ — экспонент орграфа Γ;
- (i, j) — дуга в орграфе, инцидентная вершинам i и j;
- $w(i_0, i_1, \ldots, i_k)$ — путь в орграфе, последовательно проходящий через вершины $i_0, i_1, \ldots, i_k$, $k \in \mathbb{N}$;
- $w[i, j]$ — путь $w(i_0, i_1, \ldots, i_k)$, где $i = i_0$ и $j = i_k$;
- $c(i_0, i_1, \ldots, i_k)$ — контур в орграфе, последовательно проходящий через вершины $i_0, i_1, \ldots, i_k$, $k \in \mathbb{N}$;
- $w \cdot w'$ — конкатенация путей w и w' в орграфе (определена, если и только если конечная вершина пути w совпадает с начальной вершиной пути w');
- $tC(i)$ — t-кратно пройденный контур C, начиная из вершины i;
- $\operatorname{len} w$ ($\operatorname{len} c$) — длина пути w (контура c), равная числу дуг пути (контура);
- $\langle L \rangle$ — аддитивная полугруппа, порождённая множеством L, где $L \subset \mathbb{N}$;
- МАГ — модифицированный аддитивный генератор.

Точное определение существенных переменных для степеней преобразования $\varphi : X^n \to X^n$, где X — векторное пространство над конечным полем, связано, как правило, с кратным просмотром таблиц координатных функций, т. е. вычислительная сложность задачи зависит от n экспоненциально. Поэтому для исследования перемешивающих свойств таких преобразований применяется оценочный матрично-графовый подход, при реализации которого существенная зависимость координат выходных векторов от координат входных кодируется 0,1-матрицей $M(\varphi) = (m_{ij})$ порядка n: $m_{ij} = 1 \Leftrightarrow i \in E(\varphi_j), i, j \in \{0, \ldots, n-1\}$, где $\varphi_0, \ldots, \varphi_{n-1}$ — координатные функции преобразования φ. Матрица $M(\varphi)$ называется перемешивающей матрицей преобразования φ. Равносильно для исследования перемешивающих свойств рассматривается n-вершинный перемешивающий орграф $\Gamma(\varphi)$, матрица смежности вершин которого совпадает с $M(\varphi)$. Важными характеристиками перемешивающей матрицы (орграфа) является примитивность, т. е. положительность матрицы $M(\varphi)$ в некоторой степени, и экспонент примитивной матрицы $M(\varphi)$ (орграфа $\Gamma(\varphi)$): $\exp \Gamma(\varphi) = \exp M(\varphi) = \min\{\gamma \in \mathbb{N} : M(\varphi)^\gamma > 0\}$. Значение экспонента есть нижняя оценка числа итераций преобразования, после которых достигается полное перемешивание входных данных, т. е. зависимость каждой выходной координаты от всех входных координат.

В работе исследуется примитивность перемешивающих орграфов преобразований некоторых классов регистров сдвига длины n над множеством V_r с двумя обратными связями, оцениваются их экспоненты. Полученные выводы развивают результаты работ [1, 2] для регистров сдвига длины n над V_r с одной обратной связью.

В п. 1 доказан критерий биективности регистровых преобразований с произвольным числом обратных связей. В п. 2 исследованы множества простых путей и контуров перемешивающего орграфа регистра с двумя обратными связями. В п. 3 описан частный класс регистров сдвига с двумя обратными связями, построенных на основе модифицированных аддитивных генераторов. В п. 4 доказан критерий примитивности и получены оценки экспонента перемешивающего орграфа для частного класса регистров сдвига с двумя обратными связями.

1. Биективность регистровых преобразований из класса $R(n, X, t)$

При $n, r, t \in \mathbb{N}$, где $n > t \geqslant 1, r > 1$, обозначим: $R(n, X, t)$ — класс регистров сдвига длины n над множеством X с t обратными связями; $\varphi^X : X^n \to X^n$ — преобразование конечного множества X, реализуемое регистром из класса $R(n, X, t)$.

Рассмотрим функции $f_i(x_0, \ldots, x_{n-1}) : X^n \to X, i = 1, \ldots, t$, где $t < n$; X — векторное пространство над конечным полем; $x_0, \ldots, x_{n-1} \in X$. Пусть $Y \subset \{x_0, \ldots, x_{n-1}\}$, $|Y| = t$.

Система функций $F_t = \{f_1(x_0, \ldots, x_{n-1}), \ldots, f_t(x_0, \ldots, x_{n-1})\}$ называется биективной по множеству переменных Y, если F_t реализует подстановку множества X^t при любой фиксации переменных из множества $\{x_0, \ldots, x_{n-1}\} \setminus Y$.

Пусть $t > 1$ и $j_1, \ldots, j_t$ — номера ячеек регистра сдвига из класса $R(n, X, t)$, в которые записываются в каждом такте значения функций обратной связи. Преобразование $\varphi^X(x_0, \ldots, x_{n-1})$ регистра сдвига с функциями обратной связи $f_{j_s}(x_0, \ldots, x_{n-1}) : X^n \to X$, $s = 1, \ldots, t$, $j_1, \ldots, j_{t-1} \in \{0, \ldots, n-2\}$, $j_t = n-1$, определим системой координатных функций $\{\varphi_0^X(x_0, \ldots, x_{n-1}), \ldots, \varphi_{n-1}^X(x_0, \ldots, x_{n-1})\}$ в соответствии с формулами

$$\varphi_i^X(x_0,\dots,x_{n-1}) = x_{i+1} + \xi_i(x_0,\dots,x_i,x_{i+2},\dots,x_{n-1}),\ i=0,\dots,n-3; \quad (1)$$

$$\varphi_{n-2}^X(x_0,\dots,x_{n-1}) = x_{n-1} + \xi_{n-2}(x_0,\dots,x_{n-2}); \quad (2)$$

$$\varphi_{n-1}^X(x_0,\dots,x_{n-1}) = x_0 + \xi_{n-1}(x_1,\dots,x_{n-1}), \quad (3)$$

где $\xi_i(x_0,\dots,x_i,x_{i+2},\dots,x_{n-1}) = 0$ для всех $i \in \{0,\dots,n-3\} \setminus \{j_1,\dots,j_{t-1}\}$; $\xi_{n-2}(x_0, \dots,x_{n-2}) = 0$, если $j_{t-1} \neq n-2$, и $\xi_{n-1}(x_1,\dots,x_{n-1}) \neq 0$.

Обозначим: $F_t = \{f_{j_s}(x_0,\dots,x_{n-1}) : s = 1,\dots,t\}$ — система функций обратных связей регистра из класса $R(n,X,t)$; Z — множество переменных $\{x_0, x_{j_1+1},\dots,x_{j_{t-1}+1}\}$.

Теорема 1. Преобразование φ^X регистра сдвига из класса $R(n,X,t)$ биективно, если и только если система функций F_t биективна по множеству переменных Z.

Доказательство. Пусть преобразование φ^X регистра сдвига небиективно. Тогда в множестве X^n найдутся наборы $\alpha = (\alpha_0,\dots,\alpha_{n-1})$ и $\beta = (\beta_0,\dots,\beta_{n-1}), \alpha \neq \beta$, такие, что $\varphi^X(\alpha) = \varphi^X(\beta)$, то есть $\varphi_i^X(\alpha_0,\dots,\alpha_{n-1}) = \varphi_i^X(\beta_0,\dots,\beta_{n-1})$, $i = 0,\dots,n-1$. Из (1)–(3) получаем $\alpha_k = \beta_k$ для всех $k \in \{1,\dots,n-1\} \setminus \{j_1+1,\dots,j_{t-1}+1\}$. Так как $\alpha \neq \beta$, то $(\alpha_0,\alpha_{j_1+1},\dots,\alpha_{j_{t-1}+1}) \neq (\beta_0,\beta_{j_1+1},\dots,\beta_{j_{t-1}+1})$. Вместе с тем из равенства $\varphi^X(\alpha) = \varphi^X(\beta)$ следует, что $(f_{j_1}(\alpha),\dots,f_{j_t}(\alpha)) = (f_{j_1}(\beta),\dots,f_{j_t}(\beta))$. Значит, система функций F_t при некоторой фиксации переменных $\{x_0,\dots,x_{n-1}\} \setminus Z$ реализует неинъективное и, следовательно, небиективное преобразование множества X^t.

В обратную сторону. Если φ^X — подстановка, то $\varphi^X(\alpha) \neq \varphi^X(\beta)$ для любых $\alpha = (\alpha_0,\dots,\alpha_{n-1})$ и $\beta = (\beta_0,\dots,\beta_{n-1}), \alpha \neq \beta$. Возьмём α и β такие, что $\alpha_k = \beta_k$ для всех $k \in \{0,\dots,n-1\} \setminus \{0, j_1+1,\dots,j_{t-1}+1\}$. Из (1)–(3) следует, что при таких α и β неравенство $\varphi^X(\alpha) \neq \varphi^X(\beta)$ выполнено, только если $(f_{j_1}(\alpha),\dots,f_{j_t}(\alpha)) \neq$ $\neq (f_{j_1}(\beta),\dots,f_{j_t}(\beta))$. В силу произвольности выбора элементов α_k получаем для $k = \{0,\dots,n-1\} \setminus \{0, j_1+1,\dots,j_{t-1}+1\}$, что при любой фиксации переменных $\{x_0,\dots,x_{n-1}\} \setminus Z$ система функций F_t задаёт инъективное преобразование, то есть подстановку множества X^t. ■

Обозначим через $R(n,r,t)$ класс регистров сдвига $R(n,X,t)$ при $X = V_r$. Преобразование $\varphi(z_0,\dots,z_{n-1})$ из $R(n,r,t)$ есть преобразование множества $V_{nr} = \{(z_0,\dots,z_{n-1}) : z_0,\dots,z_{n-1} \in V_r\}$, где $z_k = (x_{rk},\dots,x_{r-1+rk})^{\mathrm{T}}$ есть k-й столбец двоичной матрицы, определяющей состояние регистра, $k = 0,\dots,n-1$.

2. Перемешивающие свойства регистровых подстановок из $R(n,r,2)$

Рассмотрим регистровую подстановку $\varphi(z_0,\dots,z_{n-1}) \in R(n,r,2)$ с обратными связями f_{n-1} и f_m, $0 \leqslant m < n-1$, определённую формулой

$$\varphi(z_0,\dots,z_{n-1}) = \begin{cases} (f_m, z_2,\dots,z_{n-1},f_{n-1}), & \text{если } m=0,\\ (z_1,\dots,f_m,\dots,z_{n-1},f_{n-1}), & \text{если } 0<m<n-2,\\ (z_1,\dots,z_{n-2},f_m,f_{n-1}), & \text{если } m=n-2, \end{cases} \quad (4)$$

где $z_0,\dots,z_{n-1} \in V_r$; $f_{n-1} = z_0 \oplus \psi_1$ и $f_m = z_{m+1} \oplus \psi_2$; функции $\psi_1 : V_{r(n-1)} \to V_r$ и $\psi_2 : V_{r(n-2)} \to V_r$ отличны от констант (вместо операции суммирования векторов из V_r может быть рассмотрена любая бинарная операция на V_r, биективная по обеим переменным). Заметим, что в соответствии с теоремой 1 система функций $\{f_m(z_0,\dots,z_{n-1}), f_{n-1}(z_0,\dots,z_{n-1})\}$ биективна по множеству переменных $\{z_0, z_{m+1}\}$.

Для исследования перемешивающих свойств подстановки φ используем оценочный матрично-графовый подход. В перемешивающем nr-вершинном орграфе $\Gamma(\varphi)$ подстановки φ обозначим вершины числами $u+ri$, координатные булевы функции — φ_{u+ri},

$u = 0, \ldots, r-1$, $i = 0, \ldots, n-1$. В $\Gamma(\varphi)$ пара $(v+ri, u+rj)$ есть дуга, если и только если $(v + ri) \in E(\varphi_{u+rj})$, $v, u \in \{0, \ldots, r-1\}$, $i, j \in \{0, \ldots, n-1\}$. Для описания множества дуг орграфа $\Gamma(\varphi)$ достаточно в силу (4) описать множество существенных переменных координатных функций $\varphi_{u+r(n-1)}$ и φ_{u+rm}.

Из (4) следует, что орграф $\Gamma(\varphi)$ содержит независимые простые контуры c_u длины n, $u = 0, \ldots, r-1$, где $c_u = c(u + r(n-1), u + r(n-2), \ldots, u + r, u)$.

Пусть θ — функция отождествления вершин орграфа $\Gamma(\varphi)$: при $u = 0, \ldots, r-1$ и любом $i = 0, \ldots, n-1$ положим $\theta(u + ri) = u$. Функция θ индуцирует функцию Θ, отображающую nr-вершинный орграф $\Gamma(\varphi)$ в r-вершинный орграф $\Gamma(\psi) = \Gamma(\psi_1) \cup \cup \Gamma(\psi_2)$, где пара (v, u) образует дугу орграфа $\Gamma(\psi_1)$ (орграфа $\Gamma(\psi_2)$), если и только если функция $\varphi_{u+r(n-1)}$ (функция φ_{u+rm}) зависит существенно хотя бы от одной из переменных $x_v, x_{v+r}, \ldots, x_{v+r(n-1)}$, $v, u \in \{0, \ldots, r-1\}$. На рис. 1 изображена часть перемешивающего орграфа $\Gamma(\varphi)$ подстановки φ, содержащая контуры c_v и c_u при $v \neq u$, которая при отождествлении вершин преобразуется в дугу (v, u) орграфа $\Gamma(\psi)$. Здесь координатные функции φ_{u+rm} и $\varphi_{u+r(n-1)}$ зависят существенно от некоторых из переменных $x_v, x_{v+r}, \ldots, x_{v+r(n-1)}$, $v, u \in \{0, \ldots, r-1\}$, что соответствует светлым дугам на рис. 1.

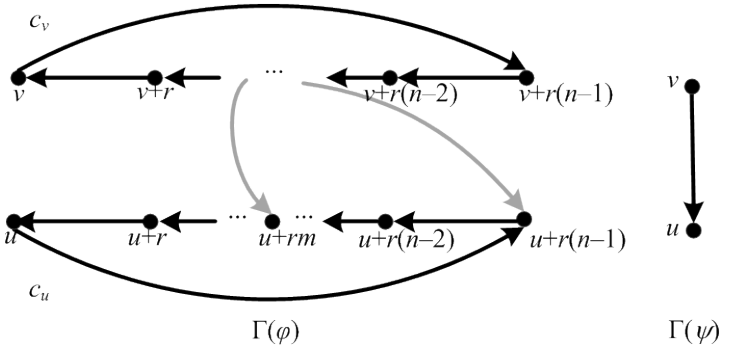

Рис. 1. Часть орграфа $\Gamma(\varphi)$ и соответствующая ей дуга в $\Gamma(\psi)$

Обозначим через Γ_0 орграф $\Gamma(\varphi)$ при функциях $\psi_1 \equiv \psi_2 \equiv 0$, то есть Γ_0 состоит из независимых простых контуров $c_0, \ldots, c_{r-1}$ длины n.

Теорема 2. Перемешивающий орграф $\Gamma(\varphi)$ сильносвязный, если и только если орграф $\Gamma(\psi)$ сильносвязный.

Доказательство. Множества вершин орграфов Γ_0 и $\Gamma(\varphi)$ совпадают и Γ_0 является частью орграфа $\Gamma(\varphi)$. Значит, если обе вершины орграфа $\Gamma(\varphi)$ принадлежат контуру c_u, то они взаимно достижимы, $u = 0, \ldots, r-1$. Следовательно, в орграфе $\Gamma(\varphi)$ при $u \neq v$, $u, v \in \{0, \ldots, r-1\}$, вершина $u+rj$ контура c_u достижима из вершины $v+ri$ контура c_v, если и только если u достижима из v в орграфе $\Gamma(\psi)$. ■

Используя индуктивный метод, опишем множество простых путей орграфа $\Gamma(\varphi)$.

При $v, u = 0, \ldots, r-1$ обозначим: $D(v, u)$ и $\Delta(v, u)$ — множества номеров переменных из множества $\{v, v + r, \ldots, v + r(n-1)\}$, существенных для координатных функций $\varphi_{u+r(n-1)}$ и φ_{u+rm} соответственно; $w[u + ri, u + rj]$ — путь в орграфе $\Gamma(\varphi)$, $i, j \in \{0, \ldots, n-1\}$. Так как простой путь $w[u+ri, u+rj]$ является частью контура c_u, его длина равна $i - j$, если $i \geqslant j$, и $n - j + i$, если $i < j$.

Пусть в орграфе $\Gamma(\psi)$ имеется дуга (v, u) (простой путь длины 1). Тогда в орграфе $\Gamma(\varphi)$ при любых $i, j \in \{0, \ldots, n-1\}$ имеются пути $w[v + ri, u + rj]$ (рис. 2) двух видов:

$$w[v+ri, v+ra]\cdot(v+ra, u+r(n-1))\cdot w[u+r(n-1), u+rj],$$
$$w[v+ri, v+rb]\cdot(v+rb, u+rm)\cdot w[u+rm, u+rj],$$

где $v+ra \in D(v,u)$, $v+rb \in \Delta(v,u)$. Заметим, что множество данных путей есть полный прообраз $\Theta^{-1}(v,u)$.

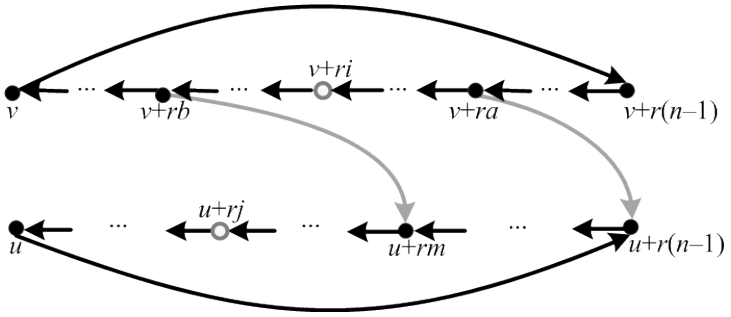

Рис. 2. Пути в орграфе $\Gamma(\varphi)$

Теперь, используя индукцию, опишем все прообразы относительно Θ для любого простого пути $w(u_1,\ldots,u_l)$ длины $l-1$ в $\Gamma(\psi)$, $l>2$. Пусть $w(u_1,\ldots,u_l) = w(u_1, \ldots, u_{l-1})\cdot(u_{l-1},u_l)$ и описаны все пути из множества $\Theta^{-1}(w(u_1,\ldots,u_{l-1}))$. В частности, описаны пути видов $w' = w[u_1{+}ri, u_{l-1}{+}ra]$ и $w'' = w[u_1{+}ri, u_{l-1}{+}rb]$ при $i=0,\ldots,n{-}1$ и любых a,b, таких, что $u_{l-1}+ra \in D(u_{l-1},u_l)$, $u_{l-1}+rb \in \Delta(u_{l-1},u_l)$. Тогда любой путь $w[u_1+ri, u_l+rj]$ из $\Theta^{-1}(w(u_1,\ldots,u_{l-1}))$ при любых $i,j \in \{0,\ldots,n-1\}$ есть конкатенация путей двух видов:

$$w[u_1+ri, u_l+rj] = w'\cdot(u_{l-1}+ra, u_l+r(n-1))\cdot w[u_l+r(n-1), u_l+rj],$$
$$w[u_1+ri, u_l+rj] = w''\cdot(u_{l-1}+rb, u_l+rm)\cdot w[u_l+rm, u_l+rj].$$

3. Регистровые преобразования на основе модифицированных аддитивных генераторов

В общем случае множества $D(v,u)$ и $\Delta(v,u)$ зависят от пары (v,u), $v,u=0,\ldots r-1$, следовательно, в орграфе $\Gamma(\varphi)$ описание путей и контуров большой длины является громоздким (см. п. 2). Опишем пути в важном для криптографических приложений случае, когда $D(v,u)=D$, $\Delta(v,u)=\Delta$ (множества $D(v,u)$ и $\Delta(v,u)$ одинаковы для всех пар (v,u)). В частности, таким свойством обладает преобразование множеств состояний аддитивных генераторов и некоторых их модификаций. Рассмотрим модификации, заключающиеся в применении преобразования к множеству значений функции обратной связи [3, 4]. При итерациях преобразования в некоторых таких модификациях достигается полное перемешивание входных данных, в то время как преобразования аддитивных генераторов плохо перемешивают входные данные. Определим аддитивные генераторы и их модификации.

Пусть b_r — биекция $\mathbb{Z}_{2^r} \leftrightarrow V_r$, определяющая двоичное r-разрядное представление числа $X \in \mathbb{Z}_{2^r}$ по правилу: если $X = 2^{r-1}x_0 + \ldots + 2x_{r-2} + x_{r-1}$, то $b_r(X) = \overline{X} = (x_0,\ldots,x_{r-1}) \in V_r$; b_r^{-1} — обратная к b_r функция. Аддитивный генератор есть регистр сдвига длины n над $\mathbb{Z}_{2^r}$ с функцией обратной связи $f: V_{nr} \to V_r$ следующего вида:

$$f(\overline{X}_0,\ldots,\overline{X}_{n-1}) = b_r\left(\left(\sum_{k\in D} X_k\right) \bmod 2^r\right),$$

где $D = \{d_0, \ldots, d_q\}$ — множество номеров существенных переменных функции f; $0 < q$; $0 = d_0 < \ldots < d_q < n$. Модифицированный аддитивный генератор (МАГ) есть регистр сдвига длины n с функцией обратной связи f^g:

$$f^g(\overline{X}_0, \ldots, \overline{X}_{n-1}) = g(f(\overline{X}_0, \ldots, \overline{X}_{n-1})) = b_r g\left(\left(\sum_{k\in D} X_k\right) \bmod 2^r\right),$$

то есть к значениям функции f применяется преобразование g множества V_r (в записи вида $b_r g(\cdot)$ умножение функций выполняется слева направо).

Преобразование φ^g множества состояний МАГ имеет вид

$$\varphi^g(\overline{X}_0, \ldots, \overline{X}_{n-1}) = (\overline{X}_1, \ldots, \overline{X}_{n-1}, f^g(\overline{X}_0, \ldots, \overline{X}_{n-1})),$$

при этом φ^g — подстановка множества V_{nr}, если и только если g — подстановка множества V_r [4, теорема 1]. Перемешивающие свойства преобразования φ^g описаны в [4].

Актуальной задачей является улучшение перемешивающих свойств регистров, построенных на основе МАГ, то есть построение преобразований, перемешивающий орграф которых имеет более низкую оценку экспонента. Один из способов решения задачи — увеличение числа обратных связей в регистровом преобразовании. Исследуем перемешивающие свойства регистровых преобразований $\varphi^{g,\mu}$ с двумя обратными связями, построенных на основе МАГ (класс таких преобразований обозначим МАГ$(n, r, 2)$), где модификация выполнена с помощью преобразований g и μ множества V_r. В соответствии с (4) регистровое преобразование $\varphi^{g,\mu}$ с использованием преобразований g и μ при $0 \leqslant m < n-1$ задано одним из следующих равенств:

$$\varphi^{g,\mu}(\overline{X}_0, \ldots, \overline{X}_{n-1}) = (f_m, \overline{X}_2, \ldots, \overline{X}_{n-1}, f_{n-1}),\ m = 0; \tag{5}$$

$$\varphi^{g,\mu}(\overline{X}_0, \ldots, \overline{X}_{n-1}) = (\overline{X}_1, \ldots, f_m, \ldots, \overline{X}_{n-1}, f_{n-1}),\ 0 < m < n-2; \tag{6}$$

$$\varphi^{g,\mu}(\overline{X}_0, \ldots, \overline{X}_{n-1}) = (\overline{X}_1, \ldots, \overline{X}_{n-2}, f_m, f_{n-1}),\ m = n-2. \tag{7}$$

Функции обратных связей f_{n-1} и f_m определены равенствами

$$f_{n-1} = b_r g\left(\left(\sum_{k\in D} X_k\right) \bmod 2^r\right); \tag{8}$$

$$f_m = b_r \mu\left(\left(\sum_{k\in \Delta} X_k\right) \bmod 2^r\right), \tag{9}$$

где $D, \Delta \subseteq \{0, \ldots, n-1\}$ — непустые множества номеров тех из чисел $X_0, \ldots, X_{n-1}$, которые суммируются в формулах (8) и (9) соответственно.

По теореме 1 преобразование $\varphi^{g,\mu}$ биективное, если и только если система функций обратной связи $\{f_m(X_0, \ldots, X_{n-1}), f_{n-1}(X_0, \ldots, X_{n-1})\}$ биективна по множеству переменных $\{X_0, X_{m+1}\}$. В частности, $\varphi^{g,\mu}$ является подстановкой, если $0 \in D \setminus \Delta$, $(m+1) \in \Delta$ и каждая координатная функция преобразований g и μ отлична от константы.

4. Примитивность и оценки экспонента перемешивающего орграфа МАГ$(n, r, 2)$

Получим критерий примитивности перемешивающего орграфа $\Gamma(\varphi^{g,\mu})$ и оценим его экспонент при $0 < m < n-2$ (при $m = n-2$ и 0 результаты получаются аналогично). Обозначим: $D = \{d_0, \ldots, d_q\}$, $\Delta = \{\delta_0, \ldots, \delta_p\}$, где $q, p > 0$; $0 = d_0 < \ldots < d_q < n$; $\delta_0 = m+1$; $0 < \delta_1 < \ldots < \delta_p < n$.

В [4] с использованием комбинаторных свойств биекции $\mathbb{Z}_{2^r} \leftrightarrow V_r$ описано множество существенных переменных функции обратной связи f^g, что позволило исследовать примитивность перемешивающего орграфа $\Gamma(\varphi^g)$ и оценить его экспонент. Описание множества существенных переменных функции f^g [4, теорема 2] справедливо для функций вида $b_r g\left(\left(\sum\limits_{k\in D} X_k\right) \bmod 2^r\right)$ при любом подмножестве $D \subseteq \{0,\dots,n-1\}$ порядка не менее 2. Воспользуемся данным описанием для исследования перемешивающих свойств регистровых преобразований $\varphi^{g,\mu}$ из МАГ$(n,r,2)$.

Опишем множество дуг перемешивающего орграфа $\Gamma(\varphi^{g,\mu})$. Обозначим при $u = 0,\dots,r-1$: $\varphi^{g,\mu}_{u+rk}(\overline{X}_0,\dots,\overline{X}_{n-1})$ — координатная булева функция преобразования $\varphi^{g,\mu}$, $k = 0,\dots,n-1$; $g_u(y_0,\dots,y_{r-1})$ и $\mu_u(y_0,\dots,y_{r-1})$ — координатные булевы функции преобразований g и μ соответственно.

Из (6) следует, что при $u = 0,\dots,r-1$ и любом $k \in \{0,\dots,n-2\}\setminus\{m\}$ существенные переменные координатных функций $\varphi^{g,\mu}_{u+rk}$ описываются просто:

$$E(\varphi^{g,\mu}_{u+rk}) = \{u + r(k+1)\}. \tag{10}$$

То есть задача состоит в описании существенных переменных функций при $k = n-1$ и $k = m$. Из (6) следует также, что в орграфе $\Gamma(\varphi^{g,\mu})$ имеется контур $c_u = (u + r(n-1), u + r(n-2),\dots,u)$, если и только если в $\Gamma(\varphi^{g,\mu})$ имеются дуги $(u + r(m+1), u + rm)$ и $(u, u + r(n-1))$, $u = 0,\dots,r-1$. Из (8), (9) следует:

$$\varphi^{g,\mu}_{u+r(n-1)}(\overline{X}_0,\dots,\overline{X}_{n-1}) = g_u\left(b_r\left(\left(\sum_{k\in D} X_k\right) \bmod 2^r\right)\right),$$
$$\varphi^{g,\mu}_{u+rm}(\overline{X}_0,\dots,\overline{X}_{n-1}) = \mu_u\left(b_r\left(\left(\sum_{k\in \Delta} X_k\right) \bmod 2^r\right)\right).$$

Обозначим $\xi(u)$ — наименьший номер существенной переменной функции $g_u(y_0,\dots,y_{r-1})$, $0 \leqslant \xi(u) < r$; $\eta(u)$ — наименьший номер существенной переменной функции $\mu_u(y_0,\dots,y_{r-1})$, $0 \leqslant \eta(u) < r$.

Из [4, теорема 2] следует описание множеств существенных переменных функций $\varphi^{g,\mu}_{u+r(n-1)}$ и $\varphi^{g,\mu}_{u+rm}$, $u = 0,\dots,r-1$.

Теорема 3. Переменная x_{v+rk} существенная:

а) для $\varphi^{g,\mu}_{u+r(n-1)}$, если и только если $k \in D$ и $\xi(u) \leqslant v < r$;

б) для $\varphi^{g,\mu}_{u+rm}$, если и только если $k \in \Delta$ и $\eta(u) \leqslant v < r$.

Данное описание множества дуг перемешивающего орграфа $\Gamma(\varphi^{g,\mu})$ позволяет исследовать условия его примитивности. Необходимым условием примитивности орграфа является сильная связность. Получим достаточное условие сильной связности орграфа $\Gamma(\varphi^{g,\mu})$.

Из (10) и теоремы 3 следует, что в орграфе $\Gamma(\varphi^{g,\mu})$ имеются пути следующих видов:

1) $w_v^{(n-1,m+1)} = w(v + r(n-1), v + r(n-2),\dots,v + r(m+1))$ и $w_v^{(m,0)} = w(v + rm, v + r(m-1),\dots,v)$ при $0 \leqslant v < \max\{\xi(u),\eta(u)\}$;
2) $w_v^{(m,m+1)} = w_v^{(m,0)} \cdot (v, v + r(n-1)) \cdot w_v^{(n-1,m+1)}$ при $v \geqslant \xi(u)$;
3) $w_v = w_v^{(n-1,m+1)} \cdot (m+1, m) \cdot w_v^{(m,0)} = w(v + r(n-1), v + r(n-2),\dots,v)$ при $v \geqslant \eta(u)$.

При $\max\{\xi(u),\eta(u)\} \leqslant v < r$ в $\Gamma(\varphi^{g,\mu})$ имеются простые контуры

$$c_v = (v + r(n-1), v + r(n-2),\dots,v).$$

Обозначим: $\Gamma(g)$ и $\Gamma(\mu)$ — перемешивающие орграфы преобразований g и μ соответственно; $V(v)$ — множество вершин $\{v + r(n-1), v + r(n-2), \ldots, v\}$ в $\Gamma(\varphi^{g,\mu})$, $v = 0, \ldots, r-1$.

Теорема 4. Орграф $\Gamma(\varphi^{g,\mu})$ сильносвязный, если в каждом из орграфов $\Gamma(g)$ и $\Gamma(\mu)$ имеется дуга $(0, r-1)$ и полустепень захода каждой вершины орграфов $\Gamma(g)$ и $\Gamma(\mu)$ больше нуля.

Доказательство. Наличие в каждом из орграфов $\Gamma(g)$ и $\Gamma(\mu)$ дуги $(0, r-1)$ равносильно тому, что переменная y_0 существенная для координатных функций $g_{r-1}(y_0, \ldots, y_{r-1})$ и $\mu_{r-1}(y_0, \ldots, y_{r-1})$. В силу теоремы 3 это равносильно тому, что переменная x_{j+rk} является существенной для функции $\varphi^{g,\mu}_{u+r(n-1)}$ при любом $k \in D$, $j = 0, \ldots, r-1$, и для функции $\varphi^{g,\mu}_{u+rm}$ при любом $k \in \Delta$, $j = 0, \ldots, r-1$. Следовательно, с использованием пути w_{r-1} и контура c_{r-1} любая вершина из $V(r-1)$ достижима из любой вершины множества $\bigcup\limits_{j=0}^{r-1} V(j)$, то есть из любой вершины орграфа $\Gamma(\varphi^{g,\mu})$.

Вместе с тем по условию в любую вершину j каждого из орграфов $\Gamma(g)$ и $\Gamma(\mu)$ заходит дуга, значит, по теореме 3 и с использованием путей вида $w_v^{(n-1,m+1)}$ и $w_v^{(m,0)}$ любая вершина орграфа $\Gamma(\varphi^{g,\mu})$ достижима из множества $V(r-1)$. Значит, орграф $\Gamma(\varphi^{g,\mu})$ сильносвязный. ■

В частности, условиям теоремы 4 удовлетворяют следующие модифицирующие преобразования:

— сдвиг $R(y_0, \ldots, y_{r-1}) = (y_1, y_2, \ldots, y_{r-1}, y_0)$ координат векторов из V_r;
— инволютивная перестановка $I(y_0, \ldots, y_{r-1}) = (y_{r-1}, \ldots, y_0)$ координат векторов из V_r;
— треугольная подстановка T множества V_r с координатными функциями $t_i(y_0, \ldots, y_{r-1}) = y_0 \oplus \ldots \oplus y_i$, $i = 0, \ldots, r-1$;
— совершенные преобразования (s-боксы множества V_r, в которых каждая координатная функция существенно зависит от всех входных переменных).

Докажем критерий примитивности орграфа $\Gamma(\varphi^{g,\mu})$. Напомним [5], что множество контуров $C = \{C_1, \ldots, C_s\}$ длин $l_1, \ldots, l_s$ соответственно называется примитивным, если числа $l_1, \ldots, l_s$ взаимно простые.

Обозначим: $d^{(\zeta)}$ — наибольшее число из D, не превышающее ζ, где $\zeta = 0, \ldots, n-1$; $\delta^{(\zeta)}$ — наибольшее число из Δ, не превышающее ζ, в случае $\delta_1 \leqslant \zeta$, где $\zeta = 1, \ldots, n-1$ в силу свойства $0 \in D \setminus \Delta$. Отсюда $d^{(n-1)} = d_q$, $\delta^{(n-1)} = \max\{\delta_0, \delta_p\}$. Пусть $d^{(m)} = d_t$, где $t \in \{0, \ldots, q\}$, и $\delta^{(m)} = \delta_\tau$ при $\delta_1 \leqslant m$, где $\tau \in \{0, \ldots, p\}$.

Определим множества чисел L (при $\delta_1 \leqslant m$) и L' (при $\delta_1 > m$):

$$L = \{n - d_i, n + m + 1 - d_j - \delta_k, m + 1 - \delta_l : i = 0, \ldots, q,\ j = 0, \ldots, t,$$
$$k = \tau + 1, \ldots, p,\ l = 1, \ldots, \tau\},$$
$$L' = \{n - d_i, n + m + 1 - d_j - \delta_k : i = 0, \ldots, q,\ j = 0, \ldots, t,\ k = 1, \ldots, p\}.$$

Теорема 5 (критерий примитивности орграфа $\Gamma(\varphi^{g,\mu})$). Сильносвязный орграф $\Gamma(\varphi^{g,\mu})$ примитивный, если и только если либо $\delta_1 \leqslant m$ и L — множество взаимно простых чисел, либо $\delta_1 > m$ и L' — множество взаимно простых чисел.

Доказательство. Напомним, что в соответствии с универсальным критерием примитивности [6, ч. 1, разд. 11.3] сильносвязный орграф примитивный, если и только если он содержит примитивную систему простых контуров (иначе говоря, длины всех простых контуров образуют множество взаимно простых чисел).

Обозначим: $B_i = \{v + ri : v = 0, \ldots, r-1\}$ — множество вершин, $i = 0, \ldots, n-1$; $C^{(v)}$ — множество простых контуров сильносвязного орграфа $\Gamma(\varphi^{g,\mu})$, все вершины которых принадлежат множеству $V(v)$ вершин контура c_v, где $v \in \{0, \ldots, r-1\}$. В соответствии с теоремой 3 $C^{(v)} \neq \varnothing$, если $\max\{\xi(u), \eta(u)\} \leqslant v < r$ (в частности, $C^{(v)} \neq \varnothing$, если отлична от константы каждая координатная функция преобразований g и μ).

Из (6) следует, что каждый контур орграфа $\Gamma(\varphi^{g,\mu})$ проходит хотя бы через одну из вершин множества $B_{n-1} \cup B_m$. В частности, каждый контур из $C^{(v)}$ проходит хотя бы через одну из двух вершин $v + r(n-1)$ и $v + rm$. Из (8) и (9) следует, что $C^{(v)}$ есть объединение трёх множеств:

$$C^{(v)} = \{c_v^D(d) : d \in D\} \cup \{c_v^{\Delta,D}(\delta, d) : d \in D, \delta \in \Delta, d \leqslant m < \delta\} \cup \{c_v^{\Delta}(\delta) : \delta \in \Delta, \delta \leqslant m\},$$

где

$$\begin{gathered} c_v^D(d) = w(v + r(n-1), v + r(n-2), \ldots, v + rd) \cdot (v + rd, v + r(n-1)), \\ c_v^{\Delta}(\delta) = w(v + rm, v + r(m-1), \ldots, v + r\delta) \cdot (v + r\delta, v + rm), \\ c_v^{\Delta,D}(\delta, d) = w(v + r(n-1), v + r(n-2), \ldots, v + r\delta) \cdot (v + r\delta, v + rm) \cdot \\ \cdot w(v + rm, v + r(m-1), \ldots, v + rd) \cdot (v + rd, v + r(n-1)). \end{gathered}$$

Длины этих контуров равны

$$\operatorname{len} c_v^D(d) = n - d, \ \operatorname{len} c_v^{\Delta}(\delta) = m + 1 - \delta, \ \operatorname{len} c_v^{\Delta,D}(\delta, d) = n + m + 1 - \delta - d.$$

Следовательно, множество длин контуров из $C^{(v)}$ совпадает при $\delta_1 \leqslant m$ с множеством L и при $\delta_1 > m$ — с множеством L'. Тогда в соответствии с универсальным критерием примитивности взаимная простота либо множества L при $\delta_1 \leqslant m$, либо множества L' при $\delta_1 > m$ достаточна для примитивности сильносвязного орграфа $\Gamma(\varphi^{g,\mu})$.

Докажем необходимость. Пусть $d, d' \in D$; $\delta, \delta' \in \Delta$.

С л у ч а й 1. $\delta_1 \leqslant m$.

В $\Gamma(\varphi^{g,\mu})$ при $v, u \in \{0, \ldots, r-1\}$ элементарным путём назовём любой путь вида

$$w_{v,u}^{(n-1,n-1)}(d) = w(v + r(n-1), \ldots, v + rd) \cdot (v + rd, u + r(n-1)), \ m + 1 \leqslant d < n; \quad (11)$$

$$w_{v,u}^{(m,m)}(\delta) = w(v + rm, \ldots, v + r\delta) \cdot (v + r\delta, u + rm), \ 0 \leqslant \delta \leqslant m; \quad (12)$$

$$w_{v,u}^{(n-1,m)}(\delta) = w(v + r(n-1), \ldots, v + r\delta) \cdot (v + r\delta, u + rm), \ m + 1 \leqslant \delta < n; \quad (13)$$

$$w_{v,u}^{(m,n-1)}(d) = w(v + rm, \ldots, v + rd) \cdot (v + rd, u + r(n-1)), \ 0 \leqslant d \leqslant m. \quad (14)$$

Элементарные пути $w_{v,u}^{(n-1,n-1)}(d)$, $w_{v,u}^{(m,m)}(\delta)$, $w_{v,u}^{(n-1,m)}(\delta)$ и $w_{v,u}^{(m,n-1)}(d)$ изображены светлым цветом на рис. 3 и 4. Заметим, что при $v = u$ элементарные пути вида (11) и (12) являются контурами.

Каждый контур орграфа $\Gamma(\varphi^{g,\mu})$ проходит через некоторую вершину множества $B_{n-1} \cup B_m$, поэтому он однозначно представляется конкатенацией элементарных путей при фиксированной начальной вершине из множества $B_{n-1} \cup B_m$. Рангом контура в орграфе $\Gamma(\varphi^{g,\mu})$ назовём число составляющих его элементарных путей.

Докажем с помощью индукции по рангу контура промежуточное утверждение: длина любого контура орграфа $\Gamma(\varphi^{g,\mu})$ содержится в полугруппе $\langle L \rangle$.

Если утверждение верно и орграф $\Gamma(\varphi^{g,\mu})$ примитивный, то числа множества L взаимно простые, иначе у длин всех контуров орграфа $\Gamma(\varphi^{g,\mu})$ имеется общий делитель.

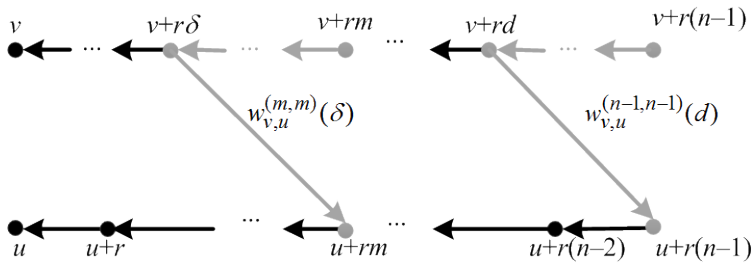

Рис. 3. Элементарные пути $w_{v,u}^{(n-1,n-1)}(d)$ и $w_{v,u}^{(m,m)}(\delta)$

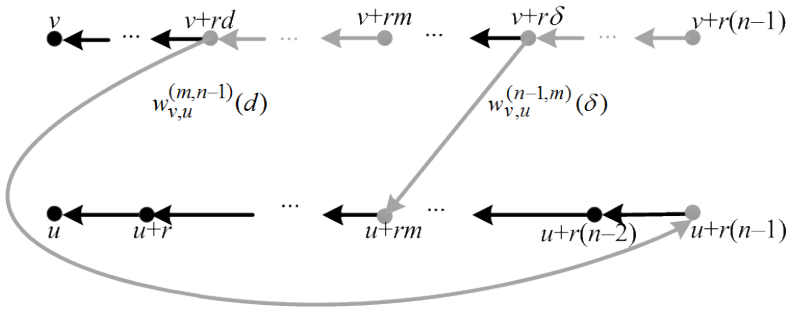

Рис. 4. Элементарные пути $w_{v,u}^{(n-1,m)}(\delta)$ и $w_{v,u}^{(m,n-1)}(d)$

В соответствии с теоремой 3 любой контур ранга 1 имеет при $v = u$ вид (11), где $\xi(u) \leqslant v < r$, или вид (12), где $\eta(u) \leqslant v < r$. Длина его равна соответственно $n - d$ и $m + 1 - \delta$. Следовательно, для простых контуров ранга 1 утверждение верно.

Любой контур ранга 2 есть конкатенация двух элементарных путей и при $v, u \in \{0, \ldots, r-1\}$ имеет вид либо $w_{v,u}^{(n-1,n-1)}(d) \cdot w_{u,v}^{(n-1,n-1)}(d')$, если $m + 1 \leqslant d < n$ и $m + 1 \leqslant d' < n$, либо $w_{v,u}^{(m,m)}(\delta) \cdot w_{u,v}^{(m,m)}(\delta')$, если $0 \leqslant \delta \leqslant m$ и $0 \leqslant \delta' \leqslant m$, либо $w_{v,u}^{(n-1,m)}(\delta) \cdot w_{u,v}^{(m,n-1)}(d)$ или $w_{v,u}^{(m,n-1)}(d) \cdot w_{u,v}^{(n-1,m)}(\delta)$, если $0 \leqslant d \leqslant m$ и $m + 1 \leqslant \delta < n$. Длины таких контуров равны $n - d + n - d'$, $m + 1 - \delta + m + 1 - \delta'$ и $n + m + 1 - d - \delta$ соответственно. Следовательно, для контуров ранга 2 утверждение также верно.

Пусть утверждение верно для всех контуров рангов $k-1$ и $k-2$, где $k > 2$. Докажем, что утверждение верно для любого контура ранга k.

Любой контур ранга k является конкатенацией k элементарных путей и при $v, u, z \in \{0, \ldots, r-1\}$ имеет одно из следующих строений:

$$w_1 \cdot w_{u,v}^{(n-1,n-1)}(d), \text{ если } m + 1 \leqslant d < n;$$
$$w_2 \cdot w_{u,v}^{(m,m)}(\delta), \text{ если } 0 \leqslant \delta \leqslant m;$$
$$w_3 \cdot w_{z,u}^{(n-1,m)}(\delta) \cdot w_{u,v}^{(m,n-1)}(d),$$
$$w_4 \cdot w_{z,u}^{(m,n-1)}(d) \cdot w_{u,v}^{(n-1,m)}(\delta), \text{ если } 0 \leqslant d \leqslant m \text{ и } m + 1 \leqslant \delta < n,$$

где $w_1 = w[v+r(n-1), u+r(n-1)]$, $w_2 = w[v+rm, u+rm]$, $w_3 = w[v+r(n-1), z+r(n-1)]$, $w_4 = w[v + rm, z + rm]$ — пути в $\Gamma(\varphi^{g,\mu})$. В соответствии с равенствами (11)–(14)

$$\operatorname{len} w_{u,v}^{(n-1,n-1)}(d) = n - d, \quad \operatorname{len} w_{u,v}^{(m,m)}(\delta) = m + 1 - \delta,$$
$$\operatorname{len}(w_{z,u}^{(n-1,m)}(\delta) \cdot w_{u,v}^{(m,n-1)}(d)) = \operatorname{len}(w_{z,u}^{(m,n-1)}(d) \cdot w_{u,v}^{(n-1,m)}(\delta)) = n + m + 1 - d - \delta.$$

Следовательно, осталось показать, что длины путей w_1, w_2, w_3 и w_4 содержатся в $\langle L\rangle$. Длины путей w_1 и w_2 совпадают с длинами контуров ранга $k-1$, полученными из этих путей заменой последних дуг. Длины путей w_3 и w_4 совпадают с длинами контуров ранга $k-2$, полученными из этих путей заменой последних дуг. По предположению индукции длины контуров рангов $k-1$ и $k-2$ содержатся в $\langle L\rangle$, значит, длины путей w_1, w_2, w_3 и w_4 также содержатся в $\langle L\rangle$. Следовательно, длина любого контура ранга k содержится в $\langle L\rangle$.

С л у ч а й 2. $\delta_1 > m$.

В этом случае в $\Gamma(\varphi^{g,\mu})$ имеются элементарные пути трёх видов $w_{v,u}^{(n-1,n-1)}(d)$, $w_{v,u}^{(n-1,m)(\delta)}$ и $w_{v,u}^{(m,n-1)}(d)$, которые определены равенствами (11), (13) и (14) соответственно.

С помощью индукции по рангу контура докажем промежуточное утверждение: длина любого контура орграфа $\Gamma(\varphi^{g,\mu})$ содержится в полугруппе $\langle L'\rangle$. В этом случае из примитивности орграфа $\Gamma(\varphi^{g,\mu})$ следует взаимная простота чисел множества L'.

В соответствии с теоремой 3 любой контур ранга 1 имеет вид (11) при $v = u$, где $\xi(u) \leqslant v < r$. Длина его равна $n - d$. Следовательно, для простых контуров ранга 1 утверждение верно.

Любой контур ранга 2 есть конкатенация двух элементарных путей и при $v, u \in \{0, \ldots, r-1\}$ имеет вид либо $w_{v,u}^{(n-1,n-1)}(d) \cdot w_{u,v}^{(n-1,n-1)}(d')$, если $m+1 \leqslant d < n$ и $m+1 \leqslant d' < n$, либо $w_{v,u}^{(n-1,m)}(\delta) \cdot w_{u,v}^{(m,n-1)}(d)$ или $w_{v,u}^{(m,n-1)}(d) \cdot w_{u,v}^{(n-1,m)}(\delta)$, если $0 \leqslant d \leqslant m$ и $m+1 \leqslant \delta < n$. Длины таких контуров равны $n-d+n-d'$ и $n+m+1-d-\delta$ соответственно. Отсюда утверждение верно для контуров ранга 2.

Пусть утверждение верно для всех контуров рангов $k-1$ и $k-2$, где $k > 2$. Докажем, что утверждение верно для любого контура ранга k. Любой контур ранга k является конкатенацией k элементарных путей и при $v, u, z \in \{0, \ldots, r-1\}$ имеет одно из следующих строений:

$$w_1 \cdot w_{u,v}^{(n-1,n-1)}(d), \text{ если } m+1 \leqslant d < n;$$
$$w_2 \cdot w_{z,u}^{(n-1,m)}(\delta) \cdot w_{u,v}^{(m,n-1)}(d),$$
$$w_3 \cdot w_{z,u}^{(m,n-1)}(d) \cdot w_{u,v}^{(n-1,m)}(\delta), \text{ если } 0 \leqslant d \leqslant m \text{ и } m+1 \leqslant \delta < n,$$

где $w_1 = w[v+r(n-1), u+r(n-1)]$, $w_2 = w[v+r(n-1), z+r(n-1)]$, $w_3 = w[v+rm, z+rm]$ — пути в $\Gamma(\varphi^{g,\mu})$. В соответствии с равенствами (11), (13), (14)

$$\operatorname{len} w_{u,v}^{(n-1,n-1)}(d) = n-d,$$
$$\operatorname{len}(w_{z,u}^{(n-1,m)}(\delta) \cdot w_{u,v}^{(m,n-1)}(d)) = \operatorname{len}(w_{z,u}^{(m,n-1)}(d) \cdot w_{u,v}^{(n-1,m)}(\delta)) = n+m+1-d-\delta.$$

Следовательно, осталось показать, что длины путей w_1, w_2 и w_3 содержатся в $\langle L'\rangle$. Длина пути w_1 равна длине контура ранга $k-1$, полученного из w_1 заменой последней дуги. Длины путей w_2 и w_3 равны длинам контуров ранга $k-2$, полученным из этих путей заменой последних дуг. По предположению индукции длины контуров рангов $k-1$ и $k-2$ содержатся в $\langle L'\rangle$, следовательно, длины путей w_1, w_2 и w_3 также содержатся в $\langle L'\rangle$. Следовательно, длина любого контура ранга k содержится в $\langle L'\rangle$. Теорема доказана. ■

Получим оценки экспонента орграфа $\Gamma(\varphi^{g,\mu})$. Обозначим:

— $\rho(D) = \max\{n-d_q, d_q-d_{q-1}, \ldots, d_1-d_0\}$ при $d_q > m$;
— $\rho(\Delta, D) = \max\{\max\{n-\delta_p, \delta_p-\delta_{p-1}, \ldots, \delta_{\tau+2}-\delta_0\}+m-d_q+1, \max\{m-d_q+1, d_q-d_{q-1}, \ldots, d_1-d_0\}\}$ при $d_q \leqslant m$;

— $\rho(\Delta) = \max\{\delta_1 + n - \delta_p, \delta_p - \delta_{p-1}, \ldots, \delta_0 - \delta_\tau, \ldots, \delta_2 - \delta_1\}$ при $\delta_1 \leqslant m$;
— $\rho'(\Delta, D) = \max\{\max\{n - \delta_p, \delta_p - \delta_{p-1}, \ldots, \delta_1 - \delta_0\}, \max\{m - d_t + 1, d_t - d_{t-1}, \ldots, d_1 - d_0\} + n - \delta_p\}$ при $\delta_1 > m$;
— $\varepsilon = \max\{2n - m - 2 - d_q, n + m - \max\{\delta_0, \delta_p\}\}$;
— $\varepsilon' = \max\{2m + 1 - \delta_\tau, n - 1 - d_t\}$.

Лемма 1. Пусть $0 \leqslant v < r$, $0 \leqslant u < r$, $i, j \in \{0, \ldots, n-1\}$. Тогда в орграфе $\Gamma(\phi^{g,\mu})$ длины кратчайших путей:

а) из вершины $v + ri$ в вершину $nr - 1$ — не превышает $\rho(D)$ при $d_q > m$; $\rho(\Delta, D)$ при $d_q < m$;
б) из вершины $v + ri$ в вершину $r - 1 + rm$ — не превышает $\rho(\Delta)$ при $\delta_1 \leqslant m$; $\rho'(\Delta, D)$ при $d_q < m$;
в) из вершины $nr - 1$ в вершину $u + rj$ — не превышает ε;
г) из вершины $r - 1 + rm$ в вершину $u + rj$ — не превышает ε'.

Доказательство.

а) При $d_q > m$ и любом v кратчайший путь $w[v + ri, nr - 1]$ имеет строение

$$w[v + ri, nr - 1] = w(v + ri, v + r(i-1), \ldots, v + rd^{(i)}) \cdot (v + rd^{(i)}, nr - 1),$$

его длина равна $i - d^{(i)} + 1$. Отсюда $\operatorname{len}[v + ri, nr - 1] \leqslant \rho(D)$. При $d_q \leqslant m$ кратчайший путь $w[v + ri, nr - 1]$ имеет одно из двух строений:

$$w(v + ri, v + r(i-1), \ldots, v + r\delta^{(i)}) \cdot (v + r\delta^{(i)}, r - 1 + rm) \cdot w^{(m,n-1)}_{r-1,r-1}(d^{(m)}) \text{ при } m < i \leqslant n - 1,$$
$$w(v + ri, v + r(i-1), \ldots, v + rd^{(i)}) \cdot (v + rd^{(i)}, nr - 1) \text{ при } 0 \leqslant i \leqslant m,$$

где элементарный путь $w^{(m,n-1)}_{r-1,r-1}(d^{(m)})$ определён равенством (14). Длина пути $w[v + ri, nr - 1]$ равна $i - \delta^{(i)} + 1 + m - d^m$ в первом случае и $i - d^{(i)} + 1$ во втором. С учётом равенства $d^{(m)} = d_t = d_q$ получаем

$$\max_{i \in \{0,\ldots,n-1\}} \operatorname{len} w[v + ri, nr - 1] \leqslant \rho(\Delta, D).$$

б) При $\delta_1 \leqslant m$ кратчайший путь $w[v + ri, r - 1 + rm]$ имеет строение

$$w[v + ri, r - 1 + rm] = w(v + ri, v + r(i-1), \ldots, v + r\delta^{(i)}) \cdot (v + r\delta^{(i)}, r - 1 + rm),$$

его длина равна $i - \delta^{(i)} + 1$. Отсюда

$$\max_{i \in \{0,\ldots,n-1\}} \operatorname{len} w[v + ri, r - 1 + rm] \leqslant \rho(\Delta).$$

При $\delta_1 > m$ кратчайший путь $w[v + ri, r - 1 + rm]$ имеет одно из двух строений:

$$w(v + ri, v + r(i-1), \ldots, v + r\delta^{(i)}) \cdot (v + r\delta^{(i)}, r - 1 + rm) \text{ при } m < i \leqslant n - 1,$$
$$w(v + ri, v + r(i-1), \ldots, v + rd^{(i)}) \cdot (v + rd^{(i)}, nr - 1) \cdot w^{(n-1,m)}_{r-1,r-1}(\delta^{(n-1)}) \text{ при } 0 \leqslant i \leqslant m,$$

где элементарный путь $w^{(n-1,m)}_{r-1,r-1}(\delta^{(n-1)})$ определён равенством (13). Длина пути $w[v + ri, r - 1 + rm]$ в первом случае равна $i - \delta^{(i)} + 1$ и во втором случае равна $i - d^{(i)} + 1 + n - \delta^{(n-1)}$. С учётом равенства $\delta^{(n-1)} = d_p$ получаем

$$\max_{i \in \{0,\ldots,n-1\}} \operatorname{len} w[v + ri, r - 1 + rm] \leqslant \rho'(\Delta, D).$$

в) Кратчайший путь $w[nr-1, u+rj]$ имеет одно из двух строений:

$$w[nr-1,u+rj] = w_{r-1,u}^{(n-1,n-1)}(d^{(n-1)}) \cdot w(u+r(n-1), u+r(n-2), \ldots, u+rj)$$
$$\text{при } m < j \leqslant n-1,$$
$$w[nr-1,u+rj] = w_{r-1,u}^{(n-1,m)}(\delta^{(n-1)}) \cdot w(u+rm, u+r(m-1), \ldots, u+rj) \text{ при } 0 \leqslant j \leqslant m,$$

где элементарные пути $w_{r-1,u}^{(n-1,n-1)}(d^{(n-1)})$ и $w_{r-1,u}^{(n-1,m)}(\delta^{(n-1)})$ определены равенствами (11) и (13) соответственно. Длина пути $w[nr-1, u+rj]$ равна $2n-1-d^{(n-1)}-j$ при $m < j \leqslant n-1$ и $n+m-\delta^{(n-1)}-j$ при $0 \leqslant j \leqslant m$. С учётом равенств $d^{(n-1)} = d_q$ и $\delta^{(n-1)} = \max\{\delta_0, \delta_p\}$ получаем

$$\max_{j\in\{0,\ldots n-1\}} \operatorname{len} w[nr-1, u+rj] \leqslant \varepsilon.$$

г) Кратчайший путь $w[r-1+rm, u+rj]$ имеет одно из двух строений:

$$w_{r-1,u}^{(m,m)}(\delta^{(m)}) \cdot w(u+rm, u+r(m-1), \ldots, u+rj) \text{ при } 0 \leqslant j \leqslant m,$$
$$w_{r-1,u}^{(m,n-1)}(d^{(m)}) w(u+r(n-1), u+r(n-2), \ldots, u+rj) \text{ при } m < j \leqslant n-1,$$

где элементарные пути $w_{r-1,u}^{(m,m)}(\delta^{(m)})$ и $w_{r-1,u}^{(m,n-1)}(d^{(m)})$ определены равенствами (12) и (14) соответственно. Длина пути $w[r-1+rm, u+rj]$ равна $2m+1-\delta^{(m)}-j$ при $0 \leqslant j \leqslant m$ и $n+m-d^{(m)}-j$ при $m < j \leqslant n-1$. С учётом равенств $\delta^{(m)} = \delta_\tau$ и $d^{(m)} = d_t$ получаем оценку

$$\max_{i\in\{0,\ldots,n-1\}} \operatorname{len} w[r-1+rm, u+rj] \leqslant \varepsilon'.$$

Лемма доказана. ■

Для оценки экспонента орграфа $\Gamma(\varphi^{g,\mu})$ используются числа Фробениуса. Числом Фробениуса $\Phi(\Lambda)$ называется наибольшее число, не принадлежащее аддитивной полугруппе $\langle\Lambda\rangle$, где Λ — множество взаимно простых чисел; $\Phi(\{1\}) = \Phi(\{1, l_2, \ldots, l_s\}) = -1$ при любых $l_2, \ldots, l_s$. Известно, что при $s = 2$ выполнено равенство $\Phi(\{l_1, l_2\}) = l_1 l_2 - l_1 - l_2$.

Теорема 6. Пусть $C = \{C_1, \ldots, C_s\}$ — примитивное множество контуров в орграфе $\Gamma(\varphi^{g,\mu})$ с множеством длин $\Lambda = \{l_1, \ldots, l_s\}$, где $s \geqslant 1$ и все вершины контуров $C_1, \ldots, C_s$ принадлежат множеству $V(r-1)$. Тогда:

а) если контуры $C_1, \ldots, C_s$ проходят через вершину $nr-1$, то

$$\exp\Gamma(\varphi^{g,\mu}) \leqslant \Phi(\Lambda) + 1 + \rho(d_q) + \varepsilon, \tag{15}$$

где $\rho(d_q) = \rho(D)$ при $d_q > m$ и $\rho(d_q) = \rho(\Delta, D)$ при $d_q \leqslant m$;

б) если контуры $C_1, \ldots, C_s$ не проходят через вершину $nr-1$, то

$$\exp\Gamma(\varphi^{g,\mu}) \leqslant \Phi(\Lambda) + 1 + \rho(\Delta) + \varepsilon'; \tag{16}$$

в) если контуры $C_1, \ldots, C_h$ проходят через вершину $nr-1$, $1 \leqslant h < s$, а контуры $C_{h+1}, \ldots, C_s$ проходят через вершину $r-1+rm$ и не проходят через вершину $nr-1$, то

$$\exp\Gamma(\varphi^{g,\mu}) \leqslant \Phi(\Lambda) + 1 + \rho(d_q) + n - \max\{\delta_0, \delta_p\} + \varepsilon'. \tag{17}$$

Доказательство. Для получения оценок экспонента примитивного орграфа $\Gamma(\varphi^{g,\mu})$ используем утверждение $3, б$ из [7, с. 103]: если в сильносвязном орграфе Γ при некотором $l \in \mathbb{N}$ имеется путь длины l из любой вершины в любую, то орграф Γ примитивный и $\exp\Gamma \leqslant l$.

Возьмём в $\Gamma(\varphi^{g,\mu})$ любые вершины $v+ri$ и $u+rj$, $0 \leqslant v < r$, $0 \leqslant u < r$, $i, j \in \in \{0, \ldots, n-1\}$, и оценим длину пути $w[v+ri, u+rj]$, используя подход, изложенный в [5] для получения универсальной оценки экспонента. Для этого рассмотрим следующие случаи.

а) Пусть контуры $C_1, \ldots, C_s$ проходят через вершину $nr-1$. Путь $w[v+ri, u+rj]$ представляется конкатенацией путей

$$w[v+ri, u+rj] = w[v+ri, nr-1] \cdot t_1 C_1(nr-1) \cdot \ldots \cdot t_s C_s(nr-1) \cdot w[nr-1, u+rj],$$

где $t_1, \ldots, t_s$ — целые неотрицательные коэффициенты, равные кратностям обхода контуров $C_1, \ldots, C_s$ соответственно. Варьируя кратности $t_1, \ldots, t_s$, можно (в соответствии с определением числа Фробениуса) построить путь $w[v+ri, u+rj]$ длины t при любом $t > \Phi(\Lambda) + \operatorname{len} w[v+ri, nr-1] + \operatorname{len} w[nr-1, u+rj]$. Отсюда

$$\exp\Gamma(\varphi^{g,\mu}) \leqslant \Phi(\Lambda) + 1 + \max_{i\in\{0,\ldots,n-1\}} \operatorname{len} w[v+ri, nr-1] + \max_{j\in\{0,\ldots,n-1\}} \operatorname{len}[nr-1, u+rj].$$

В соответствии с леммой 1 (случаи *а* и *в*) получаем оценку

$$\exp\Gamma(\varphi^{g,\mu}) \leqslant \Phi(\Lambda) + 1 + \rho(d_q) + \varepsilon,$$

где $\rho(d_q) = \rho(D)$ при $d_q > m$ и $\rho(d_q) = \rho(\Delta, D)$ при $d_q \leqslant m$.

б) Пусть контуры $C_1, \ldots, C_s$ не проходят через вершину $nr-1$, то есть $\delta_1 \leqslant m$. Путь $w[v+ri, u+rj]$ представляется конкатенацией путей

$$w[v+ri, u+rj] = w[v+ri, r-1+rm] \cdot t_1 C_1(r-1+rm) \cdot \ldots \cdot t_s C_s(r-1+rm) \cdot w[r-1+rm, u+rj],$$

где $t_1, \ldots, t_s$ — целые неотрицательные коэффициенты, равные кратностям обхода контуров $C_1, \ldots, C_s$ соответственно. Варьируя кратности $t_1, \ldots, t_s$, можно (в соответствии с определением числа Фробениуса) построить путь $w[v+ri, u+rj]$ длины t при любом $t > \Phi(\Lambda) + \operatorname{len} w[v+ri, r-1+rm] + \operatorname{len} w[r-1+rm, u+rj]$. Отсюда

$$\exp\Gamma(\varphi^{g,\mu}) \leqslant \Phi(\Lambda) + 1 + \max_{i\in\{0,\ldots,n-1\}} \operatorname{len} w[v+ri, r-1+rm] + \max_{j\in\{0,\ldots,n-1\}} \operatorname{len}[r-1+rm, u+rj].$$

В соответствии с леммой 1 (случаи *б* и *г*) получаем оценку

$$\exp\Gamma(\varphi^{g,\mu}) \leqslant \Phi(\Lambda) + 1 + \rho(\Delta) + \varepsilon'.$$

в) Пусть $C_1, \ldots, C_h$ — контуры, проходящие через вершину $nr-1$, а $C_{h+1}, \ldots, C_s$ — контуры, проходящие через вершину $r-1+rm$ и не проходящие через вершину $nr-1$. Путь $w[v+ri, u+rj]$ представляется конкатенацией путей

$$\begin{gathered} w[v+ri, u+rj] = w[v+ri, nr-1] \cdot t_1 C_1(nr-1) \cdot \ldots \cdot t_h C_h(nr-1) \cdot \\ \cdot w(r-1+n(r-1), r-1+n(r-2), \ldots, r-1+rm) \cdot t_{h+1} C_{h+1}(r-1+rm) \cdot \ldots \\ \ldots \cdot t_s C_s(r-1+rm) \cdot w[r-1+rm, u+rj], \end{gathered}$$

где $t_1, \ldots, t_s$ — целые неотрицательные коэффициенты, определяющие кратности обхода контуров $C_1, \ldots, C_s$ соответственно. Варьируя кратности $t_1, \ldots, t_s$, можно (в соответствии с определением числа Фробениуса) построить путь $w[v + ri, u + rj]$ длины t при любом $t > \Phi(\Lambda) + \operatorname{len} w[v+ri, nr-1] + \operatorname{len} w[nr-1, r-1+rm] + \operatorname{len} w[r-1+rm, u+rj]$. Отсюда

$$\exp\Gamma(\varphi^{g,\mu}) \leqslant \Phi(\Lambda) + 1 + \max_{i\in\{0,\ldots,n-1\}} \operatorname{len} w[v + ri, nr - 1] + n - \max\{\delta_0, \delta_p\} + \\ + \max_{j\in\{0,\ldots,n-1\}} \operatorname{len}[r - 1 + rm, u + rj].$$

В соответствии с леммой 1 (случаи *а* и *г*) получаем оценку

$$\exp\Gamma(\varphi^{g,\mu}) \leqslant \Phi(\Lambda) + 1 + \rho(d_q) + n - \max\{\delta_0, \delta_p\} + \varepsilon'.$$

Теорема доказана. ■

Из теоремы 6 следует, что оценка $\exp\Gamma(\varphi^{g,\mu})$ зависит от выбора параметров, в частности от значения параметра m второй обратной связи и точек съёма: $d^{(n-1)} = d_q$ и $d^{(m)} = d_t$ из множества D, $\delta^{(m)} = \delta_\tau$ и $\delta^{(n-1)} = \max\{\delta_0, \delta_p\}$ из множества Δ. Известно, что оценка экспонента понижается при наличии петель в примитивном орграфе. Уточним оценку $\exp\Gamma(\varphi^{g,\mu})$ для случая, когда $(n-1) \in D$, $m \in \Delta$.

Следствие 1. Если выполнено условие $(n-1) \in D$, $m \in \Delta$, то сильносвязный орграф $\Gamma(\varphi^{g,\mu})$ является примитивным и справедлива оценка

$$\exp\Gamma(\varphi^{g,\mu}) \leqslant \min\{\rho(D) + \varepsilon, \rho(\Delta) + \varepsilon'\}. \tag{18}$$

Доказательство. Если $(n-1) \in D$, $m \in \Delta$, то в орграфе $\Gamma(\varphi^{g,\mu})$ есть петли в вершинах $nr - 1$ и $r - 1 + rm$, следовательно, $\Gamma(\varphi^{g,\mu})$ примитивен и $\Phi(\Lambda) = -1$. Так как $d_q = n - 1$, $\delta_\tau = m$, то выполняются условия $d_q > m$, $\delta_1 \leqslant m$, следовательно, $\rho(d_q) = \rho(D)$. Из случаев *а* и *б* теоремы 6 получаем оценку следствия. ■

В [4, теорема 5] получены оценки экспонента перемешивающего орграфа $\Gamma(\varphi^g)$ регистрового преобразования на основе МАГ с одной обратной связью:

$$\exp\Gamma(\varphi^g) \leqslant \Phi(\Lambda) + \rho(D) + 2n - d_q; \tag{19}$$

$$\exp\Gamma(\varphi^g) \leqslant \rho(D) + n, \text{ при} (n-1) \in D. \tag{20}$$

Проиллюстрируем полученные результаты на примерах.

Пример 1.

а) Рассмотрим перемешивающий орграф $\Gamma(\varphi^g)$ регистра сдвига с одной обратной связью при $n = 8$, $r = 32$, $D = \{0, 2, 4, 5\}$. Оценим $\exp\Gamma(\varphi^g)$. В орграфе $\Gamma(\varphi^g)$ множество длин контуров $L = \{8, 6, 4, 3\}$ (рис. 5). Орграф $\Gamma(\varphi^g)$ примитивный, так как НОД$(8, 6, 4, 3) = 1$. Выберем примитивное подмножество контуров с множеством длин $\Lambda = \{3, 4\}$. С учётом равенств $\Phi(3, 4) = 5$, $\rho(D) = 3$, $d_q = 5$ получаем по формуле (19) оценку $\exp\Gamma(\varphi^g) \leqslant 19$.

б) Оценим $\exp\Gamma(\varphi^{g,\mu})$ перемешивающего орграфа $\Gamma(\varphi^{g,\mu})$ регистра с двумя обратными связями при тех же значениях n, r, D. Пусть $m = 3$, $\Delta = \{2, 4, 5\}$. В этом случае

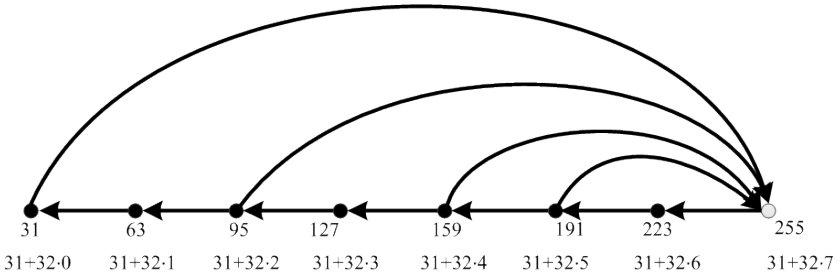

Рис. 5. Часть орграфа $\Gamma(\varphi^{g})$ при $v = 31$

выполняются условия $d_q > m$ ($d_q = 5$) и $\delta_1 \leqslant m$ ($\delta_1 = 2$). В $\Gamma(\varphi^{g,\mu})$ имеется множество $C^{(31)}$ (рис. 6), состоящее из семи простых контуров:

$$c^{D}_{31}(0) = c(255, 223, 191, 159, 127, 95, 63, 31),\ c^{D}_{31}(2) = c(255, 223, 191, 159, 127, 95),$$
$$c^{D}_{31}(4) = c(255, 223, 191, 159),\ c^{D}_{31}(5) = c(255, 223, 191),$$
$$c^{\Delta,D}_{31}(5, 0) = (255, 223, 191, 127, 95, 63, 31),$$
$$c^{\Delta,D}_{31}(5, 2) = (255, 223, 191, 127, 95),\ c^{\Delta}_{31}(2) = (127, 95).$$

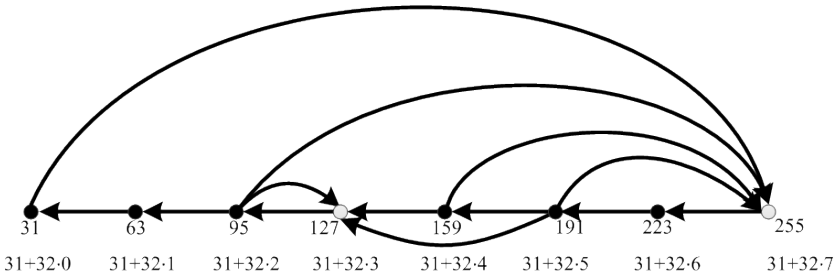

Рис. 6. Часть орграфа $\Gamma(\varphi^{g,\mu})$ при $v = 31$

Множество длин данных контуров есть $L = \{8, 7, 5, 4, 3, 2\}$. В соответствии с теоремой 5 орграф $\Gamma(\varphi^{g,\mu})$ примитивный. В соответствии с теоремой 6 оценим $\exp\Gamma(\varphi^{g,\mu})$ в зависимости от используемого примитивного множества контуров C. Заметим, что $d_t = \delta_\tau = 2$, $\max\{\delta_0, \delta_p\} = 5$. Значения параметров, необходимых для оценки экспонента, даны в табл. 1.

Таблица 1

Значения параметров в зависимости от выбора C

Параметры	Значения параметров (теорема 6, случай *а*)	Значения параметров (теорема 6, случай *в*)
C	$C = \{c^{D}_{31}(5), c^{D}_{31}(4)\}$	$C = \{c^{D}_{31}(5), c^{\Delta}_{31}(2)\}$
Λ, $\Phi(\Lambda)$	$\Lambda = \{3, 4\}$, $\Phi(3, 4) = 5$	$\Lambda = \{2, 3\}$, $\Phi(2, 3) = 1$
$\rho(d_q)$	$\rho(d_q) = \rho(D) = 3$	$\rho(d_q) = \rho(D) = 3$
ε либо ε'	$\varepsilon = 6$	$\varepsilon' = 5$

В случае *а* теоремы 6 (формула (15)) $\exp\Gamma(\varphi^{g,\mu}) \leqslant 15$; в случае *в* теоремы 6 (формула (17)) $\exp\Gamma(\varphi^{g,\mu}) \leqslant 13$.

Таким образом, оценка экспонента перемешивающего орграфа $\Gamma(\varphi^{g,\mu})$ регистра с двумя обратными связями может быть понижена по сравнению с оценкой экспонента

перемешивающего орграфа $\Gamma(\varphi^g)$ регистра с одной обратной связью. Величина оценки $\exp\Gamma(\varphi^{g,\mu})$ может быть оптимизирована с помощью выбора примитивного множества контуров.

в) Сравним (табл. 2) при $n = 8$ и $r = 32$ оценку $\exp\Gamma(\varphi^{g,\mu})$ по формулам (15) и (17) с известными оценками [6, ч. 1, разд. 11.3], применёнными к примитивным nr-вершинным орграфам Γ. Абсолютная оценка Виландта имеет вид

$$\exp\Gamma \leqslant (nr)^2 - 2nr + 2.$$

При известной длине l контура в примитивном орграфе Γ оценка более точная:

$$\exp\Gamma \leqslant nr + l(nr - 2).$$

Оценка экспонента уточняется, если в орграфе известны длины l и λ двух простых контуров, где $(l, \lambda) = 1$, $1 < \lambda < l \leqslant nr$. Если эти контуры не имеют общих вершин, то

$$\exp\Gamma \leqslant l\lambda - 2l - 3\lambda + 3nr;$$

если контуры имеют h общих вершин, то

$$\exp\Gamma \leqslant l\lambda - l - 3\lambda + h + 2nr.$$

Таблица 2

Оценки $\exp\Gamma(\varphi^{g,\mu})$, вычисленные по разным формулам

Формула	Значение оценки
$(nr)^2 - 2nr + 2$	65026
$nr + l(nr - 2)$	764 при $l = 2$
$l\lambda - 2l - 3\lambda + 3nr$	762 при $\lambda = 2,\ l = 3$
$l\lambda - l - 3\lambda + h + 2nr$	511 при $\lambda = 3,\ l = 4,\ h = 2$
$\Phi(\Lambda) + 1 + \rho(d_q) + \varepsilon$	15
$\Phi(\Lambda) + 1 + n - \max\{\delta_0, \delta_p\} + \rho(d_q) + \varepsilon'$	13

Табл. 2 показывает, что оценка $\exp\Gamma(\varphi^{g,\mu})$ существенно понижена с использованием полученных в теореме 6 формул.

Пример 2. Рассмотрим перемешивающий орграф $\Gamma(\varphi^{g,\mu})$, содержащий петли (случай $(n-1) \in D$, $m \in \Delta$), при $n = 8$, $r = 32$, $D = \{0, 2, 7\}$, $\Delta = \{m, m+1, 7\}$. В этом случае $\rho(D) = 5$, $d_q = 7$, $\max\{\delta_0, \delta_p\} = 7$. Сравним (табл. 3) оценки $\exp\Gamma(\varphi^g)$ (формула (20)) и $\exp\Gamma(\varphi^{g,\mu})$ (формула (18)) при различных значениях параметра m второй обратной связи.

Таблица 3

Оценки $\exp\Gamma(\varphi^g)$ и $\exp\Gamma(\varphi^{g,\mu})$ при различных значениях параметров

m	$\rho(D)$	$\rho(\delta)$	d_t	δ_τ	ε	ε'	$\exp\Gamma(\varphi^g)$	$\exp\Gamma(\varphi^{g,\mu})$
1	5	5	0	1	6	7	13	11
2	5	4	2	2	5	5	13	9
3	5	4	2	3	4	5	13	9
4	5	5	2	4	5	5	13	10
5	5	6	2	5	6	6	13	11

Пример показывает, что экспонент перемешивающего орграфа $\Gamma(\varphi^{g,\mu})$ может иметь более низкую оценку по сравнению с экспонентом перемешивающего орграфа $\Gamma(\varphi^g)$. Наименьшая величина оценки $\exp\Gamma(\varphi^{g,\mu})$ получается при значениях $m \approx \lceil (n-2)/2 \rceil$.

Замечание 1. Полученные оценки (15)–(18) являются достижимыми при оптимальном выборе примитивной системы контуров. Доказательство достижимости получается, например, при $d_q > m$, если рассмотреть пару вершин орграфа $(v+ri, u+rj)$ при $i = d_{k-1}$, где $\max\{n-d_q, d_q-d_{q-1}, \ldots, d_1-d_0\}$ достигается при $q = k$, и при вершине $u+rj$, наиболее удалённой от вершины $u+r(n-1)$ или вершины $u+rm$ (в зависимости от соотношения параметров). В частности, в примере 1, *б* между вершинами 254 и 158 существуют пути длины t при любом $t \geqslant 13$ и в силу свойств чисел Фробениуса не существует пути длины 12. При этом получена оценка $\exp\Gamma(\varphi^{g,\mu}) \leqslant 13$.

Заключение

Установлен ряд свойств преобразований регистров сдвига над пространтсвом V_r двоичных r-мерных векторов. Для регистровых преобразований с произвольным числом обратных связей получен критерий биективности.

С использованием матрично-графового подхода исследованы перемешивающие свойства преобразований регистров сдвига над V_r. Для подстановок φ из класса регистров сдвига с двумя обратными связями описаны множества простых путей и контуров в перемешивающем орграфе $\Gamma(\varphi)$. Для перемешивающих орграфов $\Gamma(\varphi^{g,\mu})$ подстановок регистров сдвига с двумя обратными связями, построенных на основе модифицированных аддитивных генераторов, доказан критерий примитивности, получены достижимые верхние оценки экспонента.

Полученные оценки экспонентов улучшают все другие известные оценки экспонентов для тех же орграфов (табл. 2).

Показано, что оценка экспонента перемешивающего орграфа $\Gamma(\varphi^{g,\mu})$ регистра сдвига над V_r с двумя обратными связями при определённых параметрах может быть улучшена по сравнению с оценкой экспонента перемешивающего орграфа $\Gamma(\varphi^g)$ регистра сдвига с одной обратной связью. Примеры показывают, что при одинаковых множествах точек съёма и $m \approx \lceil (n-2)/2 \rceil$ оценка экспонента улучшается до 30 %. Наименьшая величина оценки $\exp\Gamma(\varphi^{g,\mu})$ достигается в случае, если $(n-1)$ и m являются точками съёма, то есть орграф $\Gamma(\varphi^{g,\mu})$ имеет петли.

Полученные результаты могут быть использованы для построения итеративных криптографических алгоритмов на основе МАГ с быстрым перемешиванием входных данных.

ЛИТЕРАТУРА

1. *Коренева А. М., Фомичёв В. М.* Об одном обобщении блочных шифров Фейстеля // Прикладная дискретная математика. 2012. № 3(17). С. 34–40.
2. *Коренева* (*Дорохова*) *А. М., Фомичёв В. М.* Уточнённые оценки экспонентов перемешивающих графов биективных регистров сдвига над множеством двоичных векторов // Прикладная дискретная математика. 2014. № 1(23). С. 77–83.
3. *Коренева* (*Дорохова*) *А. М.* Оценки экспонентов перемешивающих графов некоторых модификаций аддитивных генераторов // Прикладная дискретная математика. Приложение. 2014. № 7. С. 60–64.
4. *Коренева А. М., Фомичёв В. М.* Перемешивающие свойства модифицированных аддитивных генераторов // Дискрет. анализ и исслед. операций. 2017. Т. 24. № 2. С. 32–52.
5. *Фомичёв В. М.* Новая универсальная оценка экспонентов графов // Прикладная дискретная математика. 2016. № 3(33). С. 78–84.
6. *Фомичёв В. М., Мельников Д. А.* Криптографические методы защиты информации: учебник для академического бакалавриата. М.: ЮРАЙТ, 2016. 454 с.

7. *Фомичёв В. М.* Оценки экспонентов примитивных графов // Прикладная дискретная математика. 2011. № 2(12). С. 101–112.

REFERENCES

1. *Koreneva A. M. and Fomichev V. M.* Ob odnom obobshchenii blochnykh shifrov Feystelya [About a Feistel block cipher generalization]. Prikladnaya Diskretnaya Matematika, 2012, no. 3(17), pp. 34–40. (in Russian)
2. *Dorokhova A. M. and Fomichev V. M.* Utochnennye otsenki eksponentov peremeshivayushchikh grafov biektivnykh registrov sdviga nad mnozhestvom dvoichnykh vektorov [Improvement of exponent estimates for mixing graphs of bijective shift registers over a set of binary vectors]. Prikladnaya Diskretnaya Matematika, 2014, no. 1(23), pp. 77–83. (in Russian)
3. *Dorokhova A. M.* Otsenki eksponentov peremeshivayushchikh grafov nekotorykh modifikatsiy additivnykh generatorov [Estimates for exponents of mixing graphs relating to some modifications of additive generators]. Prikladnaya Diskretnaya Matematika. Prilozhenie, 2014, no. 7, pp. 60–64. (in Russian)
4. *Koreneva A. M. and Fomichev V. M.* Peremeshivayushchie svoystva modifitsirovannykh additivnykh generatorov [The mixing properties of modified additive generators]. Diskretn. Anal. Issled. Oper., 2017, vol. 24, no. 2, pp. 32–52. (in Russian)
5. *Fomichev V. M.* Novaya universal'naya otsenka eksponentov grafov [The new universal estimation for exponents of graphs]. Prikladnaya Diskretnaya Matematika, 2016, no. 3(33), pp. 78–84. (in Russian)
6. *Fomichev V. M. and Mel'nikov D. A.* Kriptograficheskie metody zashchity informatsii: uchebnik dlya akademicheskogo bakalavriata [Cryptographic Methods of Information Security.]. Moscow, YuRAYT Publ., 2016. 454 p. (in Russian)
7. *Fomichev V. M.* Otsenki eksponentov primitivnykh grafov [The estimates of exponents for primitive graphs]. Prikladnaya Diskretnaya Matematika, 2011, no. 2(12), pp. 101–112. (in Russian)

УДК 512.54; 519.725

ОБЩАЯ АЛГЕБРАИЧЕСКАЯ СХЕМА РАСПРЕДЕЛЕНИЯ КРИПТОГРАФИЧЕСКИХ КЛЮЧЕЙ И ЕЁ КРИПТОАНАЛИЗ[1]

В. А. Романьков, А. А. Обзор

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

Показано, что многие известные схемы алгебраического открытого распределения криптографических ключей, использующие двусторонние умножения, являются частными случаями общей схемы такого вида. В большинстве случаев схемы строятся на платформах, которые являются подмножествами линейных пространств. К ним уже неоднократно применялся метод линейного разложения, разработанный первым автором. Метод позволяет вычислять распределяемые ключи без определения секретных параметров схемы, не решая лежащих в основе схем трудно разрешимых алгоритмических проблем. В работе показано, что данный метод применим к общей схеме, то есть является в определённом смысле универсальным. Общая схема выглядит следующим образом. Пусть G — алгебраическая система, на которой определена ассоциативная операция умножения, например группа, выбранная в качестве платформы. Предположим, что G является подмножеством конечномерного линейного пространства. Сначала задаётся открытое множество элементов $g_1, \dots, g_k \in G$. Затем корреспонденты, Алиса и Боб, последовательно публикуют элементы вида $\varphi_{a,b}(f)$ для $a, b \in G$, где $\varphi_{a,b}(f) = afb$, $f \in G$ и f — заданный или предварительно построенный элемент. Распределённый ключ имеет вид $K = \varphi_{a_l,b_l}(\varphi_{a_{l-1},b_{l-1}}(\dots(\varphi_{a_1,b_1}(g_i)\dots)) = a_l a_{l-1} \dots a_1 g_i b_1 \dots b_{l-1} b_l$. Предположим, Алиса выбирает параметры a, b из конечно порождённой подгруппы A группы G, Боб выбирает аналогичные параметры из конечно порождённой подгруппы B группы G, с помощью которых они конструируют преобразования вида $\varphi_{a,b}$, использованные в схеме. Тогда при некоторых естественных предположениях относительно G, A и B показывается, что любой злоумышленник может эффективно вычислить распределяемый ключ K без вычисления использованных в схеме преобразований.

Ключевые слова: *криптография, криптоанализ, распределение ключа, линейное разложение.*

DOI 10.17223/20710410/37/4

GENERAL ALGEBRAIC CRYPTOGRAPHIC KEY EXCHANGE SCHEME AND ITS CRYPTANALYSIS

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

Dostoevskii Omsk State University, Omsk, Russia

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

We show that many known schemes of the cryptographic key public exchange protocols in algebraic cryptography using two-sided multiplications are the special cases of a general scheme of this type. In most cases, such schemes are built on the platforms

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

that are subsets of some linear spaces. They have been repeatedly compromised by the linear decomposition method introduced by the first author. The method allows to compute the exchanged keys without computing any private data and, consequently, without solving the hard algorithmic problems on which the assumptions are based. Here, we show that this method can be successfully applied to the following general scheme and, thus, is a universal one. The general scheme proceeds as follows. Let G be an algebraic system with the associative multiplication, for example, a group chosen as the platform. We assume that G is a subset of a finitely dimensional linear space. First, some public elements $g_1, \ldots, g_k \in G$ are taken. Then the correspondents, Alice and Bob, sequentially publicise the elements of the form $\varphi_{a,b}(f)$ for some $a, b \in G$, where $\varphi_{a,b}(f) = afb$, $f \in G$ and f is a given or previously built element. The exchanged key has the form

$$K = \varphi_{a_l,b_l}(\varphi_{a_{l-1},b_{l-1}}(\ldots(\varphi_{a_1,b_1}(g_i)\ldots)) = a_l a_{l-1}\ldots a_1 g_i b_1 \ldots b_{l-1} b_l.$$

We suppose that Alice chooses parameters a, b in a given finitely generated subgroup A of G, and Bob picks up parameters a, b in a finitely generated subgroup B of G to construct their transformations of the form $\varphi_{a,b}$. Under some natural assumptions about G, A and B, we show that an intruder can efficiently calculate the exchanged key K without calculation of the transformations used in the scheme.

Keywords: *cryptography, cryptanalisis, key exchange, linear decomposition.*

Введение

В алгебраической криптографии известен ряд схем шифрования и распределения ключей на алгебраических системах: группах, полугруппах, кольцах или алгебрах, в которых используются двусторонние умножения на элементы соответствующих систем. Таковы схемы Андрекута [1], Ванга и др. [2], Стикеля [3], Б. и Т. Харли [4, 5], Шпильрайна — Ушакова [6], а также ряд других схем, часть из которых изложена в [7, 8]. Сюда же можно отнести схемы, применяющие сопряжения, из которых наиболее известна схема Ко и др. [9] — некоммутативный аналог алгоритма Диффи — Хеллмана [10].

Введённый первым автором в рассмотрение метод линейного разложения, теоретические основы которого изложены в [11] (см. также [12, 13]), позволяет при определённых условиях эффективно находить в перечисленных выше и в ряде других криптографических схем и протоколов передаваемые или распределяемые данные без вычисления соответствующих секретных ключей шифрования. Основным условием является наличие в используемой платформе структуры линейного пространства. Приложения метода содержатся в приведённых выше работах, а также в [14 – 16] и ряде других работ. Заметим также, что в [17] предложен метод нелинейного разложения, уже не предполагающий наличия структуры векторного пространства у платформы, на которой строится схема. Метод хорошо работает, например, на ряде схем, построенных на полициклических группах, которые сейчас часто предлагаются в качестве платформ [18 – 20].

В данной работе мы обращаем внимание на то, что перечисленные схемы алгебраического открытого распределения ключей являются частными случаями одной общей схемы. При условиях, о которых говорилось выше, метод линейного разложения для нахождения распределяемых данных применим к этой общей схеме.

1. Описание общей схемы распределения ключей

В качестве платформы выбирается алгебраическая система G: группа, полугруппа, кольцо или алгебра. Каждая такая система наделена операцией умножения «$\cdot$». Предполагаем, что эта операция ассоциативна. Любой паре элементов $a, b \in G$ отвечает преобразование $\varphi_{a,b} : G \to G$, определённое правилом $\varphi_{a,b}(g) = a \cdot g \cdot b$, $g \in G$. В дальнейшем для упрощения записи операция «$\cdot$» не указывается.

Преобразования удовлетворяют соотношению $\varphi_{a,b} \circ \varphi_{c,d} = \varphi_{ca,bd}$ и образуют полугруппу $S(G)$. При наличии в G единицы 1 существует тождественное преобразование указанного вида $\varepsilon = \varphi_{1,1}$. Если элементы a и b обратимы, то определено обратное преобразование $\varphi_{a,b}^{-1} = \varphi_{a^{-1},b^{-1}}$. Если G — группа, то $S(G)$ также группа.

Обычно в протоколах указываются подмножества $A, B \subseteq G$, из которых выбираются элементы, участвующие в записи преобразований. При этом первый корреспондент — Алиса — выбирает эти элементы из A, а второй корреспондент — Боб — из B. Часто считается, что элементы из A и B попарно перестановочны между собой, то есть для любого $a \in A$ и любого $b \in B$ выполнено равенство $ab = ba$.

Рассматриваемые схемы устроены следующим образом. Сначала публикуются, то есть объявляются открытыми, несколько (обычно один) элементов $g_1, \ldots, g_k \in G$. Затем Алиса и Боб последовательно публикуют значения вида $\varphi_{a,b}(f)$, где $f \in G$ — уже опубликованный ранее элемент. Распределённый ключ выглядит как

$$K = \varphi_{a_l,b_l}(\varphi_{a_{l-1},b_{l-1}}(\ldots(\varphi_{a_1,b_1}(g_i)\ldots)) = a_l a_{l-1} \ldots a_1 g_i b_1 \ldots b_{l-1} b_l. \tag{1}$$

Отсюда видно, что для получения ключа K достаточно применить к элементу g_i преобразование $\varphi_{u,v}$, где $u = a_l a_{l-1} \ldots a_1$, $v = b_1 \ldots b_{l-1} b_l$. Предполагается, что все преобразования вида φ_{a_j,b_j} были использованы при публикации данных. Другими словами, для любой пары a_j, b_j среди опубликованных данных есть c и d, такие, что $d = \varphi_{a_j,b_j}(c)$. Если элементы a_j и b_j обратимы, то $c = \varphi_{a_j,b_j}^{-1}(d)$, то есть можно считать, что преобразование φ_{a_j,b_j}^{-1} также использовано и может быть включено в запись (1). Именно это обстоятельство при определённых условиях на G, A и B позволяет эффективно вычислить K, не определяя элементы a_j, b_j. Об этом говорится в п. 2. Заметим, что при этом становится несущественным, кто из корреспондентов выбирает конкретные значения a_j, b_j, определяя преобразование φ_{a_j,b_j}. Заметим, что в некоторых схемах (см., например, [4, 5]), кроме значений вида $\varphi_{a,b}(f)$, иногда публикуются суммы таких значений. Для рассматриваемых схем это несущественно, так как в процессе криптографического анализа восстанавливаются слагаемые таких сумм (см. пример из п. 3).

2. Основные леммы

Предположим, что в качестве платформы распределения ключей используется группа G, являющаяся подмножеством некоторого конечномерного линейного пространства V. Два корреспондента (Алиса и Боб) договариваются об элементе $h \in G$, а также о двух конечно порождённых подгруппах A и B группы G. Предполагается, что любой из элементов $a \in A$ перестановочен с любым из элементов $b \in B$. Все эти данные считаются открытыми.

Предположим, что в данной схеме корреспонденты, начиная с элемента h, последовательно публикуют элементы вида $\varphi_{a_i,b_i}(u) = a_i u b_i$, где $a_i, b_i \in A$ (Алиса), и $\varphi_{c_j,d_j}(u) = c_j u d_j$, где $c_j, d_j \in B$ (Боб), в записях которых u — один из уже полученных до этого элементов. Распределяемый ключ имеет вид

$$K = \varphi_{f_1,g_1}^{\varepsilon_1}(\varphi_{f_2,g_2}^{\varepsilon_2}(\ldots(\varphi_{f_t,g_t}^{\varepsilon_t}(h)\ldots)),$$

где каждая пара (f_r, g_r) совпадает либо с парой вида (a_i, b_i), либо с парой вида (c_j, d_j), $\varepsilon_r \in \{\pm 1\}$.

Следующая лемма показывает, как можно строить базисы линейных подпространств пространства V, порождаемые элементами из G определённого вида. Эта лемма содержится в [11 – 13] и приводится здесь для замкнутости изложения.

Лемма 1. Пусть A — конечно порождённая подгруппа группы G, $A = \langle a_1, \ldots, a_k \rangle$, G является подмножеством конечномерного линейного пространства V над полем $\mathbb{F}$ и h — элемент группы G. Допустим, что все основные вычисления в V, то есть сложение, умножение на скаляр, решение системы линейных уравнений, эффективны. Тогда существует эффективный способ построения базиса $E = \{e_1, \ldots, e_s\}$ линейного подпространства $\mathrm{Lin}(AhA)$, порождённого в V всеми элементами вида ahb, где $a, b \in A$.

Доказательство. Рассмотрим произвольно упорядоченное, начинающееся с h множество всех элементов вида $c^\varepsilon h d^\eta$, где $\varepsilon, \eta \in \{\pm 1\}$; c, d — элементы вида a_i или 1. Назовём это множество первым списком, обозначив его L_1. Выполним следующие операции построения части $\{e_1, e_2, \ldots\}$ базиса E, являющейся базисом линейного подпространства, порождённого списком L_1:

1) Полагаем $e_1 = h$.

2) Пусть уже построены элементы $\{e_1, \ldots, e_t\}$ базиса E. Рассматриваем следующий элемент списка $c^\varepsilon h d^\eta$. Если он линейно зависим с уже выбранными элементами, он удаляется из списка. Если независим, включается в E.

Когда первый список закончится, формируется новый произвольно упорядоченный список L_2 всех элементов вида $c^\varepsilon e_j d^\nu$, как при формировании первого списка, где e_j — элемент части E, сформированной после окончания первого списка (за исключением e_1).

Далее последовательно рассматриваются элементы списка L_2 и выполняется действие 2. После окончания действий с L_2, в результате которых будет получена часть базиса E, являющаяся базисом подпространства, порождённого списками L_1 и L_2, составляется третий список по тому же правилу, и т. д.

3) Процесс построения E закончен, если при обработке очередного списка в E не добавится ни одного элемента.

Объясним это заключение. Каждый новый список состоит из элементов предыдущего списка, домноженных слева и справа на порождающие элементы из A или их обратные, допускается, что один из домножаемых элементов равен 1. Введём обозначение $X = \{a_1^{\pm 1}, \ldots, a_k^{\pm 1}, 1\}$. Тогда первый список L_1 является подмножеством в XhX, второй — L_2 — подмножеством в X^2hX^2 и т. д. Если при обработке очередного списка $L_{i+1} \subseteq X^{i+1}hX^{i+1}$ ни один из его элементов не был включен в строящийся базис E, значит, L_{i+1} принадлежит линейному подпространству, порождённому всеми предыдущими списками (соответственно всеми уже построенными элементами базиса), т. е. $X^{i+1}hX^{i+1} \subseteq \mathrm{Lin}(\cup_{j=1}^{i} X^j h X^j)$. Но тогда $X^{i+2}hX^{i+2} \subseteq X(\mathrm{Lin}(\cup_{j=1}^{i} X^j h X^j))X \subseteq$ $\subseteq \mathrm{Lin}(\cup_{j=1}^{i+1} X^j h X^j) \subseteq \mathrm{Lin}(\cup_{j=1}^{i} X^j h X^j)$. Значит, рассмотрение списка L_{i+2} также не добавит базисных элементов, и т. д. В то же время очевидно, что рассматриваемое подпространство лежит в подпространстве, порождённом всеми списками. Из приведённого рассуждения следует, что всего списков, которые могут дать вклад в строящийся базис, не больше, чем размерность всего линейного пространства V. ■

Следующая лемма является ключевой в последующем криптографическом анализе. Предполагается, что выполнены все перечисленные выше соглашения.

Лемма 2. Пусть G — группа, являющаяся подмножеством конечномерного линейного пространства V над полем $\mathbb{F}$. Предположим, что даны элементы u и $v = \varphi_{a,b}(u)$, где $a, b \in A$ являются секретными для стороннего наблюдателя.

Тогда для любого элемента вида $w = \varphi_{c,d}(u)$, где $c, d \in B$ также неизвестны (другими словами, $w \in BuB$), можно определить элемент $z = \varphi_{a,b}(w)$, используя структуру линейного пространства V.

Доказательство. По условию $v \in AuA$. Строим базис E пространства $\mathrm{Lin}(AuA)$, как это объяснено в лемме 1. Пусть $E = \{a_1ub_1, \ldots, a_rub_r\}$, $a_i, b_i \in A$. Используя алгоритм Гаусса при решении соответствующих систем линейных уравнений, получаем разложение

$$v = \sum_{i=1}^{r} \alpha_i a_i u b_i, \ \alpha_i \in \mathbb{F}.$$

Все величины, входящие в запись (1), известны. Подставим в правую часть (1) вместо u элемент w и выполним необходимые преобразования, используя поэлементную перестановочность подгрупп A и B:

$$\sum_{i=1}^{r} \alpha_i a_i w b_i = \sum_{i=1}^{r} \alpha_i a_i c u d b_i = c \left(\sum_{i=1}^{r} \alpha_i a_i u b_i \right) d = cvd = caubd = a(cud)b = awb = z.$$

Лемма доказана. ■

Замечание 1. Приведённые результаты с очевидными поправками проходят также для случая, когда G — полугруппа (с единицей или без неё), A и B — подполугруппы.

Сформулируем *мнемоническое правило* эффективного получения элемента, вытекающее из лемм 1 и 2:

$$v = \varphi_{a,b}(u) \ (a, b \in A) \ \& \ w \in BuB \ \Rightarrow \ \varphi_{a,b}(w);$$
$$v = \varphi_{c,d}(u) \ (c, d \in B) \ \& \ w \in AuA \ \Rightarrow \ \varphi_{c,d}(w).$$

Приведённая запись означает, что, вычислив по лемме 1 базис соответствующего векторного пространства и зная элементы u и v из левой части правила (его предположений), можно эффективно вычислить образ элемента w из правой части правила (его заключения).

3. Примеры нахождения распределяемого ключа

Пример 1. Опишем протокол 1 Ванга и др. из [2]. В этой работе в качестве платформы протокола распределения ключа предлагается использовать одну из групп Артина B_n, $n \in \mathbb{N}$.

По известному представлению Лоуренс — Краммера каждая из групп B_n допускает точное матричное представление ρ над некоторым полем $\mathbb{F}$. Представление ρ и поле $\mathbb{F}$ определяются в явном виде. Обратное отображение ρ^{-1} также определяется эффективно [21]. Значит, можно предположить, что платформа G описываемого ниже протокола является частью конечномерного линейного пространства V.

Алиса и Боб соглашаются относительно выбора (неабелевой) группы G и случайно выбранного элемента $h \in G$, а также двух подгрупп A и B группы G, таких, что $ab = ba$ для любой пары элементов $a \in A$ и $b \in B$. Предполагаем, что подгруппы A и B заданы конечными множествами порождающих элементов $\{a_1, \ldots, a_n\}$ и $\{b_1, \ldots, b_m\}$ соответственно. Все эти данные считаются известными.

Алгоритм распределения ключа работает следующим образом:

— Алиса выбирает четыре элемента $c_1, c_2, d_1, d_2 \in A$, вычисляет $x = d_1c_1hc_2d_2$ и пересылает по открытой сети (публикует) элемент x, предназначенный Бобу.
— Боб выбирает шесть элементов $f_1, f_2, g_1, g_2, g_3, g_4 \in B$, вычисляет $y = g_1f_1hf_2g_2$ и $w = g_3f_1xf_2g_4$, затем публикует элемент (y, w), предназначенный Алисе.
— Алиса выбирает два элемента $d_3, d_4 \in A$, вычисляет $z = d_3c_1yc_2d_4$ и $u = d_1^{-1}wd_2^{-1}$, затем публикует элемент (z, u) для Боба.
— Боб вычисляет и публикует элемент $v = g_1^{-1}zg_2^{-1}$ для Алисы.
— Алиса вычисляет ключ $K_A = d_3^{-1}vd_4^{-1} = c_1f_1hf_2c_2$.
— Боб вычисляет ключ $K_B = g_3^{-1}ug_4^{-1} = c_1f_1hf_2c_2$, равный K_A.

В результате Алиса и Боб вырабатывают общий секретный ключ $K = K_A = K_B$.

Криптографический анализ

В протоколе использованы следующие преобразования:

$$\varphi_{d_1c_1,c_2d_2},\ \varphi_{g_1f_1,f_2g_2},\ \varphi_{g_3f_1,f_2g_4},\ \varphi_{d_3c_1,c_2d_4},\ \varphi^{-1}_{d_1,d_2},\ \varphi^{-1}_{g_1,g_2}.$$

Непосредственно проверяется, что ключ K можно представить в виде

$$K = \varphi_{c_1f_1,f_2c_2}(h) = \varphi^{-1}_{d_1,d_2}(\varphi_{d_1c_1,c_2d_2}(\varphi^{-1}_{g_1,g_2}(\varphi_{g_1f_1,f_2g_2}(h)))).$$

Проследим, что все вычисления при таком представлении ключа K может проделать любой наблюдатель (не знающий параметров преобразований), используя только леммы 1 и 2.

Результат первого преобразования $y = \varphi_{g_1f_1,f_2g_2}(h)$ известен, как передаваемое по сети сообщение.

Результат второго преобразования $\varphi^{-1}_{g_1,g_2}(y)$ определяется по мнемоническому правилу

$$v = \varphi^{-1}_{g_1,g_2}(z)\ \&\ y \in AzA\ \Rightarrow\ \varphi^{-1}_{g_1,g_2}(y) = f_1hf_2.$$

Результат третьего преобразования определяется по мнемоническому правилу

$$x = \varphi_{d_1c_1,c_2d_2}(h)\ \&\ f_1hf_2 \in BhB\ \Rightarrow\ \varphi_{d_1c_1,c_2d_2}(f_1hf_2) = d_1c_1f_1hf_2c_2d_2.$$

Результат четвертого преобразования определяется по мнемоническому правилу

$$u = \varphi^{-1}_{d_1,d_2}(w)\ \&\ d_1c_1f_1hf_2c_2d_2 \in BwB\ \Rightarrow\ \varphi^{-1}_{d_1,d_2}(d_1c_1f_1hf_2c_2d_2) = c_1f_1hf_2c_2 = K.$$

Пример 2. Известный протокол Ко и др. [9] называют *некоммутативным аналогом протокола Диффи — Хеллмана*. Предлагаемая его авторами платформа — одна из групп кос Артина B_n, $n \in \mathbb{N}$. Относительно её представления матрицами см. пример 1. Соглашения о первоначальном выборе платформы $G = B_n$, элемента $h \in G$, двух конечно порождённых подгрупп A и B группы G те же самые, что и в примере 1.

Алгоритм распределения ключа работает следующим образом:

— Алиса выбирает элемент $a \in A$, вычисляет $h^a = aha^{-1}$ и пересылает по открытой сети (публикует) элемент h^a, предназначенный Бобу.
— Боб выбирает элемент $b \in B$, вычисляет $h^b = bhb^{-1}$ и пересылает по открытой сети (публикует) элемент h^b, предназначенный Алисе.
— Алиса вычисляет ключ $K_A = (h^b)^a = h^{ab}$.
— Боб вычисляет ключ $K_B = (h^a)^b = h^{ba}$.

Так как $ab = ba$, эти ключи совпадают, предоставляя корреспондентам Алисе и Бобу секретный распределённый ключ $K = K_A = K_B$.

Криптографический анализ

Заметим, что

$$K = abha^{-1}b^{-1} = \varphi_{a,a^{-1}}(\varphi_{b,b^{-1}}(h)).$$

Результат первого преобразования $h^b = \varphi_{b,b^{-1}}(h)$ известен, как передаваемое по сети сообщение.

Результат второго преобразования определяется по мнемоническому правилу

$$h^a = \varphi_{a,a^{-1}}(h) \ \&\ h^b \in BhB \ \Rightarrow \ \varphi_{a,a^{-1}}(h^b) = K.$$

Пример 3. Опишем протокол Б. и Т. Харли [4, 5]. Пусть G — конечно порождённая коммутативная подгруппа общей линейной группы $\mathrm{GL}_n(\mathbb{F})$ над полем $\mathbb{F}$. Эти данные открыты.

Алгоритм работает следующим образом:

— Боб выбирает $y \in \mathbb{F}^n$ и элемент $b \in G$, вычисляет и публикует yb.
— Алиса хочет послать Бобу секрет $x \in \mathbb{F}^n$. Для этого она выбирает элементы $a_1, a \in G$, вычисляет и пересылает по сети сообщение (xa, yba_1), предназначенное Бобу.
— Боб выбирает $b_1, b_2 \in G$, вычисляет и пересылает сообщение (xab_1, ya_1b_2), предназначенное Алисе.
— Алиса вычисляет (xb_1, yb_2) и пересылает сообщение $xb_1 - yb_2$ для Боба.
— Боб вычисляет $x - yb_2b_1^{-1}$ и восстанавливает x.

Боб может использовать yb в дальнейшем.

Криптографический анализ

Так как платформа G — коммутативная группа, можно считать, что корреспонденты используют произвольные односторонние (правые) умножения $\rho_c = \varphi_{1,c}$, $c \in G$, а также, что $A = B = G$. Аргументы леммы 1 позволяют построить базисы подпространств вида $\mathrm{Lin}(gG)$, $g \in G$. Условие $w \in BuB$ леммы 2 выполняется автоматически. Соответственно упрощается мнемоническое правило. Подгруппа G порождает в $\mathrm{GL}_n(\mathbb{F})$ конечномерную подалгебру, на которую продолжаются действия преобразований ρ_c. Отсюда следует, что для вычисления секрета достаточно найти его выражение как элемента этой подалгебры через заданные элементы и использованные в протоколе преобразования вида ρ_c.

В рассматриваемом протоколе заданы элементы yb, xa и $xb_1 - yb_2$, использованы преобразования $\rho_{b_1}, \rho_{a_1}, \rho_{a_1b_2b^{-1}}$. Первое преобразование отвечает паре (xa, xab_1), второе — паре (yb, yba_1), третье — паре (yb, ya_1b_2).

Запишем искомое выражение:

$$x = \rho_{b_1}^{-1}(xb_1 - yb_2) + \rho_{b_1}^{-1}(\rho_{a_1}^{-1}(ya_1b_2)).$$

4. Оценка сложности алгоритмов в предложенном криптографическом анализе

Алгоритм, описанный в лемме 1, строит базис подпространства конечномерного линейного пространства над полем, используя метод Гаусса решения систем линейных уравнений. В каждом рассматриваемом случае нужно определить только наличие или отсутствие решения у данной системы. Для верхней оценки числа полевых операций

в алгоритме заметим, что это число в алгоритме Гаусса для матрицы размера $(t \times s)$ оценивается как $\mathrm{O}(t^2 s)$. Пусть r — размерность пространства V, что совпадает с числом уравнений в системах. Число неизвестных не превышает r, так как оно равно числу уже включённых в базис подпространства элементов и не может превышать размерности всего пространства. Значит, каждый раз выполняется не более $\mathrm{O}(r^3)$ операций. Общее число просматриваемых списков не превышает r, так как из каждого из них, за исключением последнего, хотя бы один элемент должен добавляться в базис. В то же время первый список должен дать хотя бы два элемента базиса. В каждом из списков не более $4k^2 r$ элементов. Всего таких элементов не более $2k^2 r^2$. Итоговая оценка даёт $\mathrm{O}(k^2 r^5)$ операций. Эта оценка весьма грубая.

В некоторых случаях необходимо учитывать предварительное представление платформы матрицами и операции, связанные с обратным переходом относительно таких представлений, чтобы записать полученный ключ в исходном виде. Например, в [21] показано, что из матричного вида элемента группы кос B_n его стандартная запись восстанавливается с использованием не более чем $\mathrm{O}(n^3 \log_2 d_t)$ операций, где d_t — некоторый эффективно вычислимый параметр, зависящий от соответствующего элемента. Заметим, что получение матричной записи элемента производится фактически за время, определяемое линейной функцией его длины.

В алгоритме леммы 2 также применяется алгоритм Гаусса. В данном случае он каждый раз находит единственное решение. Так как всего используется ограниченное число построений базиса и разложений по нему, общая оценка остается прежней: $\mathrm{O}(k^2 r^5)$.

ЛИТЕРАТУРА

1. *Andrecut M.* A Matrix Public Key Cryptosystem. arXiv math.: 1506.00277v1 [cs.CR], 31 May 2015.
2. *Wang X., Xu C., Li G., et al.* Double Shielded Public Key Cryptosystems. Cryptology ePrint Archive, Report 2014/558, Version 20140718:185200, 2014. `https://eprint.iacr.org/2014/558`
3. *Stickel E.* A new method for exchanging secret keys // Proc. Third Intern. Conf. ICITA 05. Contemp. Math. 2005. V. 2. P. 426–430.
4. *Harley B. and Harley T.* Group Ring Cryptography. arXiv: 1104.17.24v1 [math.GR], 9 April 2011. 20 p.
5. *Harley T.* Cryptographic Schemes, Key Exchange, Public Key. arXiv math.: 1305.4063v1 [cs.CR], May 2013. 19 p.
6. *Shpilrain V. and Ushakov A.* A new key exchange protocol based on the decomposition problem // Algebraic Methods in Cryptography. Contemp. Math. 2006. V. 418. P 161–167.
7. *Myasnikov A., Shpilrain V., and Ushakov A.* Group-Based Cryptography. (Advances courses in Math., CRM, Barselona). Basel; Berlin; New York: Birkhäuser Verlag, 2008. 183 p.
8. *Myasnikov A., Shpilrain V., and Ushakov A.* Non-commutative Cryptography and Complexity of Group-Theoretic Problems. (Amer. Math. Soc. Surveys and Monographs). Providence, RI: Amer. Math. Soc., 2011. 385 p.
9. *Ko K. H., Lee S. J., Cheon J. H., et al.* New public-key cryptosystem using braid groups // CRYPTO 2000. LNCS. 2000. V. 1880. P. 166–183.
10. *Романьков В. А.* Введение в криптографию. Курс лекций. М.: Форум, 2012. 240 с.
11. *Романьков В. А.* Алгебраическая криптография. Омск: Изд-во Ом. ун-та, 2013. 135 с.

12. *Романьков В. А.* Криптографический анализ некоторых известных схем шифрования, использующих автоморфизмы // Прикладная дискретная математика. 2013. № 3(21). С. 35–51.
13. *Myasnikov A. G. and Roman'kov V. A.* A linear decomposition attack // Groups Complexity Cryptology. 2015. V. 7. P. 81–94.
14. *Roman'kov V. A. and Menshov A. V.* Cryptanalysis of Andrecut's Public Key Cryptosystem. arXiv math.: 1507.01496v1 [math.GR], 6 Jul 2015.
15. *Горнова М. Н., Кукина Е. Г., Романьков В. А.* Криптографический анализ протокола аутентификации Ушакова — Шпильрайна, основанного на проблеме бинарно скрученной сопряжённости // Прикладная дискретная математика. 2015. № 2(28). С. 46–53.
16. *Roman'kov V. A.* A polynomial time algorithm for the braid double shielded public key cryptosystems // Bulletin of the Karaganda University. Mathematics Series. 2014. No. 4(84). P. 110–115; arXiv math.:1412.5277v1 [math.GR], 17 Dec 2014.
17. *Roman'kov V. A.* A nonlinear decomposition attack // Groups Complexity Cryptology. 2017. V. 8. P. 197–207.
18. *Eick B. and Kahrobaei D.* Polycyclic Groups: A New Platform for Cryptology? arXiv math.: 0411.077v1 [math.GR], 3 Nov 2004. 7 p.
19. *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.
20. *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. 14 p.
21. *Cheon J. H. and Jun B.* A polynomial time algorithm for the Braid Diffie — Hellman Conjugacy Problem // CRYPTO-2003. LNCS. 2003. V. 2729. P. 212–225.

REFERENCES

1. *Andrecut M.* A Matrix Public Key Cryptosystem. arXiv math.: 1506.00277v1 [cs.CR], 31 May 2015.
2. *Wang X., Xu C., Li G., et al.* Double Shielded Public Key Cryptosystems. Cryptology ePrint Archive, Report 2014/558, Version 20140718:185200, 2014. `https://eprint.iacr.org/2014/558`
3. *Stickel E.* A new method for exchanging secret keys. Proc. Third Intern. Conf. ICITA 05. Contemp. Math., 2005, vol. 2, pp. 426–430.
4. *Harley B. and Harley T.* Group Ring Cryptography. arXiv: 1104.17.24v1 [math.GR], 9 April 2011. 20 p.
5. *Harley T.* Cryptographic Schemes, Key Exchange, Public Key. arXiv math.: 1305.4063v1 [cs.CR], May 2013. 19 p.
6. *Shpilrain V. and Ushakov A.* A new key exchange protocol based on the decomposition problem. Algebraic Methods in Cryptography. Contemp. Math., 2006, vol. 418. pp 161–167.
7. *Myasnikov A., Shpilrain V., and Ushakov A.* Group-Based Cryptography. (Advances courses in Math., CRM, Barselona). Basel, Berlin, New York, Birkhäuser Verlag, 2008. 183 p.
8. *Myasnikov A., Shpilrain V., and Ushakov A.* Non-commutative Cryptography and Complexity of Group-Theoretic Problems. (Amer. Math. Soc. Surveys and Monographs). Providence, RI, Amer. Math. Soc., 2011. 385 p.
9. *Ko K. H., Lee S. J., Cheon J. H., et al.* New public-key cryptosystem using braid groups. LNCS, 2000, vol. 1880, pp. 166–183.
10. *Roman'kov V. A.* Vvedenie v kriptografiyu. Kurs lektsiy [Introduction to Cryptography. Lecture Course]. Moscow, Forum Publ., 2012. 240 p. (in Russian)

11. *Roman'kov V. A.* Algebraicheskaya kriptografiya [Algebraic Cryptography]. Omsk, OmSU Publ., 2013. 135 p. (in Russian)
12. *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)
13. *Myasnikov A. G. and Roman'kov V. A.* A linear decomposition attack. Groups Complexity Cryptology, 2015, vol. 7, pp. 81–94.
14. *Roman'kov V. A. and Menshov A. V.* Cryptanalysis of Andrecut's Public Key Cryptosystem. arXiv math.: 1507.01496v1 [math.GR], 6 Jul 2015.
15. *Gornova M. N., Kukina E. G., and Roman'kov V. A.* Kriptograficheskiy analiz protokola autentifikatsii Ushakova —Shpil'rayna, osnovannogo na probleme binarno skruchennoy sopryazhennosti [Cryptographic analysis of the autentification protocol by Ushakov — Shpilrain based on the bi-twisted conjugacy problem]. Prikladnaya Diskretnaya Matematika, 2015, no. 2(28), pp. 46–53. (in Russian)
16. *Roman'kov V. A.* A polynomial time algorithm for the braid double shielded public key cryptosystems. Bulletin of the Karaganda University. Mathematics Series, 2014, no. 4(84), pp. 110–115; arXiv math.:1412.5277v1 [math.GR], 17 Dec 2014.
17. *Roman'kov V. A.* A nonlinear decomposition attack. Groups Complexity Cryptology, 2017, vol. 8, pp. 197–207.
18. *Eick B. and Kahrobaei D.* Polycyclic Groups: A New Platform for Cryptology? arXiv math.: 0411.077v1 [math.GR], 3 Nov 2004. 7 p.
19. *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.
20. *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. 14 p.
21. *Cheon J. H. and Jun B.* A polynomial time algorithm for the Braid Diffie — Hellman Conjugacy Problem. CRYPTO-2003, LNCS, 2003, vol. 2729, pp. 212–225.

МАТЕМАТИЧЕСКИЕ ОСНОВЫ НАДЁЖНОСТИ ВЫЧИСЛИТЕЛЬНЫХ И УПРАВЛЯЮЩИХ СИСТЕМ

УДК 519.718

ОЦЕНКИ НЕНАДЁЖНОСТИ СХЕМ В БАЗИСЕ РОССЕРА — ТУРКЕТТА (В P_3) ПРИ НЕИСПРАВНОСТЯХ ТИПА 0 НА ВЫХОДАХ ЭЛЕМЕНТОВ[1]

М. А. Алехина*, О. Ю. Барсукова**

*Пензенский государственный технологический университет, г. Пенза, Россия
**Пензенский государственный университет, г. Пенза, Россия

Рассматривается реализация функций трёхзначной логики схемами из ненадёжных функциональных элементов в базисе Россера — Туркетта. Предполагается, что базисные элементы подвержены неисправностям типа 0 на выходах, причём переходят в неисправные состояния независимо друг от друга с вероятностью ε $(0 < \varepsilon < 1/2)$. Получены следующие результаты: 1) любую функцию трёхзначной логики можно реализовать схемой, ненадёжность которой асимптотически (при малых ε) не больше ε; 2) для любой функции, кроме константы 0 и переменной x_i $(i \in \mathbb{N})$, такая схема является асимптотически оптимальной по надёжности и функционирует с ненадёжностью, асимптотически равной ε при малых ε; 3) функции 0, x_i можно реализовать абсолютно надёжно.

Ключевые слова: *функции трёхзначной логики, схема из функциональных элементов, ненадёжность схемы, надёжность схемы, неисправности типа 0 на выходах элементов.*

DOI 10.17223/20710410/37/5

ESTIMATIONS OF UNRELIABILITY OF CIRCUITS IN ROSSER — TURKETT BASIS (IN P_3) WITH FAULTS OF TYPE 0 AT THE OUTPUTS OF GATES

M. A. Alekhina*, O. Yu. Barsukova**

*Penza State Technological University, Penza, Russia
**Penza State University, Penza, Russia

E-mail: alekhina.marina19@yandex.ru, kuzya_7@mail.ru

This work belongs to one of the most important branches of mathematical cybernetics such as the theory of the reliability of control systems. The synthesis problem for reliable control systems is one of the main problems in discrete mathematics and mathematical cybernetics. The topicality of the research in this field is due to the importance of numerous applications arising in various sections of science and technology. We consider the realization of ternary logic functions by circuits consisting

[1]Работа поддержана грантом РФФИ № 17-01-00451.

of unreliable functional elements in Rosser — Turkett basis. We assume that all the circuit elements are exposed to faults of type 0 at their outputs and pass to fault states independently with the probability ε ($\varepsilon < 1/2$). We have obtained the following results: 1) any function of ternary logic can be realized by a circuit with unreliability that is asymptotically not more than ε for small ε; 2) for any function except the constant 0 and the variable x_i ($i \in \mathbb{N}$), such a circuit has the asymptotically optimal reliability and operates with the unreliability asymptotically equal to ε for small ε; 3) the functions 0 and x_i can be realized absolutely reliably.

Keywords: *ternary logic functions, circuit from functional gates, unreliability of a circuit, reliability of a circuit, faults of type* 0.

Введение

Работа относится к одному из важнейших разделов математической кибернетики — теории синтеза, надёжности и сложности управляющих систем. Актуальность исследований в этой области обусловлена важностью многочисленных приложений, возникающих в различных разделах науки и техники.

К числу основных модельных объектов математической теории синтеза, сложности и надёжности управляющих систем относятся схемы из ненадёжных функциональных элементов, реализующие функции k-значной логики ($k \geqslant 2$). Проблема построения оптимальных по критериям надёжности и сложности схем из ненадёжных элементов является одной из наиболее важных и в то же время трудных в теории синтеза управляющих систем. Разработка специальных методов синтеза схем из ненадёжных функциональных элементов связана, главным образом, с выбранной математической моделью неисправностей. Одна из основных моделей определяется константными неисправностями на выходах или на входах элементов. В работе рассматривается задача построения асимптотически оптимальных по надёжности схем в предположении, что функциональные элементы подвержены неисправностям типа 0 на выходах элементов.

Исторически сложилось так, что сначала исследовались инверсные неисправности функциональных элементов, реализующих булевы функции. Первые существенные математические результаты, касающиеся синтеза надёжных схем из ненадёжных элементов, получил Дж. фон Нейман [1]. Он предполагал, что элементы подвержены инверсным неисправностям, когда функциональный элемент E с приписанной ему булевой ($k = 2$) функцией $e(\tilde{x})$ в неисправном состоянии, в которое переходит с вероятностью ε ($0 < \varepsilon < 1/6$), реализует функцию $\overline{e}(\tilde{x})$. С помощью итерационного метода Дж. фон Неймана произвольную булеву функцию можно реализовать схемой, вероятность ошибки на выходе которой при любом входном наборе значений переменных не превосходит $c \cdot \varepsilon$ (c — некоторая положительная, зависящая лишь от базиса, константа), т. е. ненадёжность схемы сравнима с ненадёжностью одного элемента (такие схемы в теории надёжности управляющих систем принято называть надёжными). С ростом числа итераций сложность схемы при использовании метода Дж. фон Неймана увеличивается экспоненциально.

Любой метод синтеза схем из ненадёжных элементов характеризуется двумя важными параметрами: вероятностью ошибки на выходе схемы (ненадёжностью) и сложностью схемы. Именно оптимизации сложности схем, реализующих булевы функции, уделялось главное внимание в работах Р. Л. Добрушина, С. И. Ортюкова [2, 3], Д. Улига [4] и некоторых других авторов. Задача построения асимптотически оптимальных по надёжности схем из ненадёжных элементов, подверженных тем или иным

неисправностям, ни Дж. фон Нейманом, ни другими исследователями до появления работ М. А. Алехиной не рассматривалась.

Н. Пиппенджер [5] в классическом базисе $\{x_1\&x_2, x_1 \vee x_2, \overline{x}_1\}$ построил надёжные схемы без существенного увеличения сложности в предположении, что все элементы схемы ненадёжны — подвержены инверсным неисправностям на выходах.

С. В. Яблонский [6] рассматривал задачу синтеза надёжных схем в базисе $\{x_1\&x_2, x_1 \vee x_2, \bar{x}_1, g(x_1, x_2, x_3)\}$. Он предполагал, что элемент, реализующий функцию голосования $g(x_1, x_2, x_3) = x_1x_2 \vee x_1x_3 \vee x_2x_3$, абсолютно надёжный, а конъюнктор, дизъюнктор и инвертор — ненадёжные, подвержены произвольным неисправностям, ненадёжность каждого из них не больше ε. Доказано, что для любого ε существует алгоритм, который для каждой булевой функции строит асимптотически оптимальную по сложности схему, ненадёжность которой не больше ε.

В. В. Тарасов [7] решал задачу построения схем сколь угодно высокой надёжности (ненадёжность схемы стремится к 0). Для базисов из ненадёжных функциональных элементов с двумя входами и одним выходом найдены необходимые и достаточные условия, при которых любую булеву функцию можно реализовать схемой сколь угодно высокой надёжности.

Позднее в работах М. А. Алехиной, В. В. Чугуновой, А. В. Васина, С. М. Грабовской и некоторых других авторов решалась задача реализации булевых функций асимптотически оптимальными по надёжности схемами (неветвящимися программами) при различных неисправностях элементов.

Однако сложность решаемых задач, а следовательно, и технических устройств постоянно возрастает. Многозначная логика ($k \geqslant 3$) предоставляет более широкие возможности для разработки различных алгоритмов во многих областях, она позволяет уменьшить вычислительную сложность, размеры и число соединений в различных арифметико-логических устройствах, повысить плотность размещения элементов на схемах, найти альтернативные методы решения задач. Уже сейчас многозначная логика с успехом применяется при решении многих задач и во множестве технических разработок. Среди них различные арифметические устройства, системы искусственного интеллекта и обработки данных, обработка сложных цифровых сигналов и т. д.

В [8] описан функционально полный в P_3 базис, в котором на компромиссной основе согласованы математические и технические (МДП-техники) требования и интересы, а также рассмотрены некоторые аспекты синтеза электронных схем в этом базисе. В [9] построен функционально полный в P_4 базис, реализуемый в МОП-структурах. В [10 – 12] описаны свойства k-значных функций ($k \geqslant 3$), схемы которых можно использовать для повышения надёжности исходных схем, и изложены соответствующие методы синтеза.

Задача построения надёжных (в ряде базисов — асимптотически оптимальных по надёжности) схем при $k \geqslant 3$ решена в [10, 13 – 18] при инверсных неисправностях на выходах элементов. Константные неисправности на выходах элементов ранее не исследовались.

Цель этой работы — при $k = 3$ построить асимптотически оптимальные по надёжности схемы в базисе Россера — Туркетта при неисправностях типа 0 на выходах элементов.

1. Необходимые понятия и определения

Пусть $n \in \mathbb{N}$, $E_3 = \{0, 1, 2\}$, P_3 — множество всех функций трёхзначной логики, т. е. функций $f(x_1, \ldots, x_n) : (E_3)^n \to E_3$. Рассмотрим реализацию функций из мно-

жества P_3 схемами из ненадёжных функциональных элементов в базисе Россера — Туркетта $\{0, 1, 2, J_0(x_1), J_1(x_1), J_2(x_1), \min\{x_1, x_2\}, \max\{x_1, x_2\}\}$ ($\min\{x_1, x_2\}$ будем также обозначать через &, а $\max\{x_1, x_2\}$ — через $\vee$ [19, с. 46]).

Считаем, что схема из ненадёжных элементов реализует функцию $f(\tilde{x}^n)$ ($\tilde{x}^n = (x_1, \ldots, x_n)$), если при поступлении на её входы набора $\tilde{a}^n$ при отсутствии неисправностей в схеме на выходе появляется значение $f(\tilde{a}^n)$.

Предполагаем, что элементы схемы независимо друг от друга с вероятностью ε ($0 < \varepsilon < 1/2$) подвержены неисправностям типа 0 на выходах. Эти неисправности характеризуются тем, что на нулевых входных наборах функциональный элемент выдаёт правильное значение 0 с вероятностью 1, а на остальных наборах — значение 0 с вероятностью ε, а правильное значение — с вероятностью $1 - \varepsilon$.

Пусть схема S реализует функцию $f(\tilde{x}^n)$, $\tilde{a}^n$ — произвольный входной набор схемы S, $f(\tilde{a}^n) = \tau$. Обозначим через $P_i(S, \tilde{a}^n)$ вероятность появления значения $i \in E_3$ на выходе схемы S при входном наборе $\tilde{a}^n$, а через $P_{f(\tilde{a}^n)\neq\tau}(S, \tilde{a}^n)$ — вероятность появления ошибки на выходе схемы S при входном наборе $\tilde{a}^n$. Ясно, что $P_{f(\tilde{a}^n)\neq\tau}(S, \tilde{a}^n) = P_{\tau+1}(S, \tilde{a}^n) + P_{\tau+2}(S, \tilde{a}^n)$; в выражениях $\tau + 1$ и $\tau + 2$ сложение осуществляется по mod 3. Например, если $f(\tilde{a}^n) = 0$, то вероятность появления ошибки на этом наборе $P_{f(\tilde{a}^n)\neq 0}(S, \tilde{a}^n) = P_1(S, \tilde{a}^n) + P_2(S, \tilde{a}^n)$.

Ненадёжностью схемы S, реализующей функцию $f(\tilde{x}^n)$, будем называть число $P(S)$, равное наибольшей из вероятностей появления ошибки на выходе схемы S. *Надёжность* схемы S равна $1 - P(S)$.

Очевидно, что при неисправностях типа 0 на выходах элементов ненадёжность любого базисного элемента, кроме реализующего константу 0, равна ε, а надёжность — $1 - \varepsilon$. Элемент, реализующий константу 0, функционирует абсолютно надёжно. Ясно также, что функции x_i, $i \in \mathbb{N}$, можно реализовать абсолютно надёжно, не используя функциональных элементов.

Пусть $P_\varepsilon(f) = \inf P(S)$, где инфимум берётся по всем схемам S из ненадёжных элементов, реализующим функцию $f(\tilde{x}^n)$. Схему A, реализующую f, назовём *асимптотически оптимальной по надёжности*, если $P(A) \sim P_\varepsilon(f)$ при $\varepsilon \to 0$.

2. Рекуррентное соотношение

Для построения схемы, повышающей надёжность исходных схем, будем использовать базисные элементы $E_\&$ с функцией & и $E_\vee$ с функцией $\vee$. Функционирование элемента $E_\&$ представлено в табл. 1.

Таблица 1

x_1	x_2	$x_1 \& x_2$	$P_0(E_\&, \tilde{x}^2)$	$P_1(E_\&, \tilde{x}^2)$	$P_2(E_\&, \tilde{x}^2)$
0	0	0	1	0	0
0	1	0	1	0	0
0	2	0	1	0	0
1	0	0	1	0	0
1	1	1	ε	$1 - \varepsilon$	0
1	2	1	ε	$1 - \varepsilon$	0
2	0	0	1	0	0
2	1	1	ε	$1 - \varepsilon$	0
2	2	2	ε	0	$1 - \varepsilon$

Пусть $f(\tilde{x}^n)$ — произвольная функция из P_3, зависящая от переменных $x_1, \ldots, x_n$, и S — любая схема, реализующая $f(\tilde{x}^n)$. Возьмём два экземпляра схемы S и соединим их

выходы со входами элемента E с функцией $e \in \{\&, \vee\}$. Построенную схему обозначим через D (рис. 1).

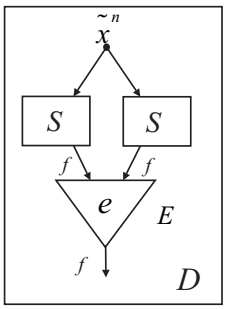

Рис. 1. Схема D

Лемма 1. Пусть $e = \&$. Тогда вероятности появления неверных значений на выходе схемы D при входном наборе $\tilde{a}^n$ удовлетворяют следующим неравенствам:

1) если $f(\tilde{a}^n) = 0$, то

$$P_1(D, \tilde{a}^n) \leqslant P_1^2(S, \tilde{a}^n) + 2P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n), \quad P_2(D, \tilde{a}^n) \leqslant P_2^2(S, \tilde{a}^n);$$

2) если $f(\tilde{a}^n) = 1$, то

$$P_0(D, \tilde{a}^n) \leqslant \varepsilon + 2P_0(S, \tilde{a}^n), \quad P_2(D, \tilde{a}^n) \leqslant P_2^2(S, \tilde{a}^n);$$

3) если $f(\tilde{a}^n) = 2$, то

$$P_0(D, \tilde{a}^n) \leqslant \varepsilon + 2P_0(S, \tilde{a}^n), \quad P_1(D, \tilde{a}^n) \leqslant 2P_1(S, \tilde{a}^n).$$

Доказательство.

1) Пусть $f(\tilde{a}^n) = 0$. Тогда правильное значение на выходе схемы D равно 0. Вычислим вероятности появления 1 и 2 на выходе схемы D, используя формулу полной вероятности:

$$\begin{gathered} P_1(D, \tilde{a}^n) = P_0^2(S, \tilde{a}^n) \cdot 0 + 2P_0(S, \tilde{a}^n)P_1(S, \tilde{a}^n) \cdot 0 + 2P_0(S, \tilde{a}^n)P_2(S, \tilde{a}^n) \cdot 0 + \\ + P_1^2(S, \tilde{a}^n)(1 - \varepsilon) + 2P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n)(1 - \varepsilon) + P_2^2(S, \tilde{a}^n) \cdot 0, \end{gathered}$$

т. е. получаем

$$P_1(D, \tilde{a}^n) = P_1^2(S, \tilde{a}^n)(1-\varepsilon) + 2P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n)(1-\varepsilon) \leqslant P_1^2(S, \tilde{a}^n) + 2P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n),$$

$$\begin{gathered} P_2(D, \tilde{a}^n) = P_0^2(S, \tilde{a}^n) \cdot 0 + 2P_0(S, \tilde{a}^n)P_1(S, \tilde{a}^n) \cdot 0 + 2P_0(S, \tilde{a}^n)P_2(S, \tilde{a}^n) \cdot 0 + \\ + P_1^2(S, \tilde{a}^n) \cdot 0 + 2P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n) \cdot 0 + P_2^2(S, \tilde{a}^n)(1 - \varepsilon), \end{gathered}$$

т. е. $P_2(D, \tilde{a}^n) = P_2^2(S, \tilde{a}^n)(1 - \varepsilon) \leqslant P_2^2(S, \tilde{a}^n)$.

2) Пусть $f(\tilde{a}^n) = 1$, тогда правильное значение на выходе схемы D равно 1. Вычислим вероятности появления 0 и 2 на выходе схемы D:

$$\begin{gathered} P_0(D, \tilde{a}^n) = P_0^2(S, \tilde{a}^n) \cdot 1 + 2P_0(S, \tilde{a}^n)P_1(S, \tilde{a}^n) \cdot 1 + 2P_0(S, \tilde{a}^n)P_2(S, \tilde{a}^n) \cdot 1 + \\ + P_1^2(S, \tilde{a}^n)\varepsilon + 2P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n)\varepsilon + P_2^2(S, \tilde{a}^n)\varepsilon. \end{gathered}$$

Принимая во внимание, что $P_1(S,\tilde{a}^n) = 1 - P_0(S,\tilde{a}^n) - P_2(S,\tilde{a}^n)$, получим

$$P_0(D,\tilde{a}^n) = P_0^2(S,\tilde{a}^n) + 2P_0(S,\tilde{a}^n)(1 - P_0(S,\tilde{a}^n) - P_2(S,\tilde{a}^n)) + 2P_0(S,\tilde{a}^n)P_2(S,\tilde{a}^n) + \\ +(1 - P_0(S,\tilde{a}^n) - P_2(S,\tilde{a}^n))^2\varepsilon + 2(1 - P_0(S,\tilde{a}^n) - P_2(S,\tilde{a}^n))P_2(S,\tilde{a}^n)\varepsilon + P_2^2(S,\tilde{a}^n)\varepsilon = \\ = 2P_0(S,\tilde{a}^n) - P_0^2(S,\tilde{a}^n) + \varepsilon - 2\varepsilon P_0(S,\tilde{a}^n) + P_0^2(S,\tilde{a}^n)\varepsilon \leqslant \varepsilon + 2P_0(S,\tilde{a}^n),$$

$$P_2(D,\tilde{a}^n) = P_0^2(S,\tilde{a}^n)\cdot 0 + 2P_0(S,\tilde{a}^n)P_1(S,\tilde{a}^n)\cdot 0 + 2P_0(S,\tilde{a}^n)P_2(S,\tilde{a}^n)\cdot 0 + P_1^2(S,\tilde{a}^n)\cdot 0 + \\ +2P_1(S,\tilde{a}^n)P_2(S,\tilde{a}^n)\cdot 0 + P_2^2(S,\tilde{a}^n)(1-\varepsilon),$$

т. е. получаем $P_2(D,\tilde{a}^n) = P_2^2(S,\tilde{a}^n)(1-\varepsilon) \leqslant P_2^2(S,\tilde{a}^n)$.

3) Пусть $f(\tilde{a}^n) = 2$, тогда правильное значение на выходе схемы D равно 2. Вычислим вероятности появления 0 и 1 на выходе схемы D:

$$P_0(D,\tilde{a}^n) = P_0^2(S,\tilde{a}^n)\cdot 1 + 2P_0(S,\tilde{a}^n)P_1(S,\tilde{a}^n)\cdot 1 + 2P_0(S,\tilde{a}^n)P_2(S,\tilde{a}^n)\cdot 1 + P_1^2(S,\tilde{a}^n)\varepsilon + \\ +2P_1(S,\tilde{a}^n)P_2(S,\tilde{a}^n)\varepsilon + P_2^2(S,\tilde{a}^n)\varepsilon.$$

Учитывая, что $P_2(S,\tilde{a}^n) = 1 - P_0(S,\tilde{a}^n) - P_1(S,\tilde{a}^n)$, получаем

$$P_0(D,\tilde{a}^n) = P_0^2(S,\tilde{a}^n) + 2P_0(S,\tilde{a}^n)P_1(S,\tilde{a}^n) + 2P_0(S,\tilde{a}^n)(1 - P_0(S,\tilde{a}^n) - P_1(S,\tilde{a}^n)) + \\ +P_1^2(S,\tilde{a}^n)\varepsilon + 2P_1(S,\tilde{a}^n)(1 - P_0(S,\tilde{a}^n) - P_1(S,\tilde{a}^n))\varepsilon + (1 - P_0(S,\tilde{a}^n) - P_1(S,\tilde{a}^n))^2\varepsilon = \\ = \varepsilon + 2P_0(S,\tilde{a}^n) + \varepsilon P_0^2(S,\tilde{a}^n) - P_0^2(S,\tilde{a}^n) - 2\varepsilon P_0(S,\tilde{a}^n) \leqslant \varepsilon + 2P_0(S,\tilde{a}^n).$$

В силу равенств $P_1(D,\tilde{a}^n) = P_1^2(S,\tilde{a}^n)(1-\varepsilon) + 2P_1(S,\tilde{a}^n)P_2(S,\tilde{a}^n)(1-\varepsilon)$ и $P_2(S,\tilde{a}^n) = \\ = 1 - P_0(S,\tilde{a}^n) - P_1(S,\tilde{a}^n)$ получаем

$$P_1(D,\tilde{a}^n) = P_1^2(S,\tilde{a}^n)(1-\varepsilon) + 2P_1(S,\tilde{a}^n)(1 - P_0(S,\tilde{a}^n) - P_1(S,\tilde{a}^n))(1-\varepsilon) = \\ = 2P_1(S,\tilde{a}^n) + {P_1}^2(S,\tilde{a}^n)\varepsilon - {P_1}^2(S,\tilde{a}^n) + 2P_1(S,\tilde{a}^n)P_0(S,\tilde{a}^n)\varepsilon - 2P_1(S,\tilde{a}^n)P_0(S,\tilde{a}^n) - \\ -2P_1(S,\tilde{a}^n)\varepsilon \leqslant 2P_1(S,\tilde{a}^n).$$

Лемма 1 доказана. ■

Функционирование базисного элемента $E_\vee$ представлено в табл. 2.

Таблица 2

x_1	x_2	$x_1 \vee x_2$	$P_0(E_\vee,\tilde{x}^2)$	$P_1(E_\vee,\tilde{x}^2)$	$P_2(E_\vee,\tilde{x}^2)$
0	0	0	1	0	0
0	1	1	ε	$1-\varepsilon$	0
0	2	2	ε	0	$1-\varepsilon$
1	0	1	ε	$1-\varepsilon$	0
1	1	1	ε	$1-\varepsilon$	0
1	2	2	ε	0	$1-\varepsilon$
2	0	2	ε	0	$1-\varepsilon$
2	1	2	ε	0	$1-\varepsilon$
2	2	2	ε	0	$1-\varepsilon$

Лемма 2. Пусть $e = \vee$. Тогда вероятности появления неверных значений на выходе схемы D (см. рис. 1) при входном наборе $\tilde{a}^n$ удовлетворяют следующим неравенствам:

1) если $f(\tilde{a}^n) = 0$, то

$$P_1(D, \tilde{a}^n) \leqslant 2P_1(S, \tilde{a}^n), \quad P_2(D, \tilde{a}^n) \leqslant 2P_2(S, \tilde{a}^n);$$

2) если $f(\tilde{a}^n) = 1$, то

$$P_0(D, \tilde{a}^n) \leqslant \varepsilon + P_0^2(S, \tilde{a}^n), \quad P_2(D, \tilde{a}^n) \leqslant 2P_2(S, \tilde{a}^n);$$

3) если $f(\tilde{a}^n) = 2$, то

$$P_0(D, \tilde{a}^n) \leqslant \varepsilon + P_0^2(S, \tilde{a}^n), \quad P_1(D, \tilde{a}^n) \leqslant P_1^2(S, \tilde{a}^n) + 2P_0(S, \tilde{a}^n)P_1(S, \tilde{a}^n).$$

Доказательство.

1) Пусть $f(\tilde{a}^n) = 0$. Тогда правильное значение на выходе схемы D равно 0. Вычислим вероятности появления 1 и 2 на выходе схемы D, используя формулу полной вероятности:

$$P_1(D, \tilde{a}^n) = 2P_0(S, \tilde{a}^n)P_1(S, \tilde{a}^n)(1-\varepsilon) + P_1^2(S, \tilde{a}^n)(1-\varepsilon).$$

Поскольку $P_0(S, \tilde{a}^n) = 1 - P_1(S, \tilde{a}^n) - P_2(S, \tilde{a}^n)$, получаем

$$\begin{gathered} P_1(D, \tilde{a}^n) = 2(1 - P_1(S, \tilde{a}^n) - P_2(S, \tilde{a}^n))P_1(S, \tilde{a}^n)(1-\varepsilon) + P_1^2(S, \tilde{a}^n)(1-\varepsilon) = \\ = 2P_1(S, \tilde{a}^n) - 2\varepsilon P_1(S, \tilde{a}^n) - P_1^2(S, \tilde{a}^n) + P_1^2(S, \tilde{a}^n)\varepsilon - 2P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n) + \\ + 2\varepsilon P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n) \leqslant 2P_1(S, \tilde{a}^n), \end{gathered}$$

$$P_2(D, \tilde{a}^n) = 2P_0(S, \tilde{a}^n)P_2(S, \tilde{a}^n)(1-\varepsilon) + 2P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n)(1-\varepsilon) + P_2^2(S, \tilde{a}^n)(1-\varepsilon).$$

Учитывая, что $P_0(S, \tilde{a}^n) = 1 - P_1(S, \tilde{a}^n) - P_2(S, \tilde{a}^n)$, получаем

$$\begin{gathered} P_2(D, \tilde{a}^n) = 2(1 - P_1(S, \tilde{a}^n) - P_2(S, \tilde{a}^n))P_2(S, \tilde{a}^n)(1-\varepsilon) + 2P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n)(1-\varepsilon) + \\ + P_2^2(S, \tilde{a}^n)(1-\varepsilon) = 2P_2(S, \tilde{a}^n) - 2P_2(S, \tilde{a}^n)\varepsilon - P_2^2(S, \tilde{a}^n) + P_2^2(S, \tilde{a}^n)\varepsilon \leqslant 2P_2(S, \tilde{a}^n). \end{gathered}$$

2) Пусть $f(\tilde{a}^n) = 1$, тогда правильное значение на выходе схемы D равно 1. Вычислим вероятности появления 0 и 2 на выходе схемы D:

$$\begin{gathered} P_0(D, \tilde{a}^n) = P_0^2(S, \tilde{a}^n) + 2P_0(S, \tilde{a}^n)P_1(S, \tilde{a}^n)\varepsilon + 2P_0(S, \tilde{a}^n)P_2(S, \tilde{a}^n)\varepsilon + P_1^2(S, \tilde{a}^n)\varepsilon + \\ + 2P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n)\varepsilon + P_2^2(S, \tilde{a}^n)\varepsilon. \end{gathered}$$

Принимая во внимание, что $P_1(S, \tilde{a}^n) = 1 - P_0(S, \tilde{a}^n) - P_2(S, \tilde{a}^n)$, получаем

$$\begin{gathered} P_0(D, \tilde{a}^n) = P_0^2(S, \tilde{a}^n) + 2P_0(S, \tilde{a}^n)(1 - P_0(S, \tilde{a}^n) - P_2(S, \tilde{a}^n))\varepsilon + 2P_0(S, \tilde{a}^n)P_2(S, \tilde{a}^n)\varepsilon + \\ + (1 - P_0(S, \tilde{a}^n) - P_2(S, \tilde{a}^n))^2\varepsilon + 2(1 - P_0(S, \tilde{a}^n) - P_2(S, \tilde{a}^n))P_2(S, \tilde{a}^n)\varepsilon + P_2^2(S, \tilde{a}^n)\varepsilon = \\ = P_0^2(S, \tilde{a}^n) - \varepsilon {P_0}^2(S, \tilde{a}^n) + \varepsilon \leqslant \varepsilon + P_0^2(S, \tilde{a}^n), \end{gathered}$$

$$P_2(D, \tilde{a}^n) = 2P_0(S, \tilde{a}^n)P_2(S, \tilde{a}^n)(1-\varepsilon) + 2P_1(S, \tilde{a}^n)P_2(S, \tilde{a}^n)(1-\varepsilon) + P_2^2(S, \tilde{a}^n)(1-\varepsilon).$$

Поскольку $P_1(S, \tilde{a}^n) = 1 - P_0(S, \tilde{a}^n) - P_2(S, \tilde{a}^n)$, получаем

$$\begin{gathered} P_2(D, \tilde{a}^n) = 2P_0(S, \tilde{a}^n)P_2(S, \tilde{a}^n)(1-\varepsilon) + 2(1 - P_0(S, \tilde{a}^n) - P_2(S, \tilde{a}^n))P_2(S, \tilde{a}^n)(1-\varepsilon) + \\ + P_2^2(S, \tilde{a}^n)(1-\varepsilon) = 2P_2(S, \tilde{a}^n) - P_2^2(S, \tilde{a}^n) + \varepsilon P_2^2(S, \tilde{a}^n) - 2\varepsilon P_2(S, \tilde{a}^n) \leqslant 2P_2(S, \tilde{a}^n). \end{gathered}$$

3) Пусть $f(\tilde{a}^n) = 2$, тогда правильное значение на выходе схемы D равно 2. Вычислим вероятности появления 0 и 1 на выходе схемы D:

$$P_0(D,\tilde{a}^n) = P_0^2(S,\tilde{a}^n) + 2P_0(S,\tilde{a}^n)P_1(S,\tilde{a}^n)\varepsilon + 2P_0(S,\tilde{a}^n)P_2(S,\tilde{a}^n)\varepsilon + P_1^2(S,\tilde{a}^n)\varepsilon + \\ +2P_1(S,\tilde{a}^n)P_2(S,\tilde{a}^n)\varepsilon + P_2^2(S,\tilde{a}^n)\varepsilon.$$

Учитывая, что $P_2(S,\tilde{a}^n) = 1 - P_0(S,\tilde{a}^n) - P_1(S,\tilde{a}^n)$, получаем

$$P_0(D,\tilde{a}^n) = P_0^2(S,\tilde{a}^n) + 2P_0(S,\tilde{a}^n)P_1(S,\tilde{a}^n)\varepsilon + 2P_0(S,\tilde{a}^n)(1 - P_0(S,\tilde{a}^n) - P_1(S,\tilde{a}^n))\varepsilon + \\ +P_1^2(S,\tilde{a}^n)\varepsilon + 2P_1(S,\tilde{a}^n)(1 - P_0(S,\tilde{a}^n) - P_1(S,\tilde{a}^n))\varepsilon + (1 - P_0(S,\tilde{a}^n) - P_1(S,\tilde{a}^n))^2\varepsilon = \\ = \varepsilon + P_0^2(S,\tilde{a}^n) - P_0^2(S,\tilde{a}^n)\varepsilon \leqslant \varepsilon + P_0^2(S,\tilde{a}^n),$$

$$P_1(D,\tilde{a}^n) = 2P_0(S,\tilde{a}^n)P_1(S,\tilde{a}^n)(1-\varepsilon) + P_1^2(S,\tilde{a}^n)(1-\varepsilon) \leqslant 2P_0(S,\tilde{a}^n)P_1(S,\tilde{a}^n) + P_1^2(S,\tilde{a}^n).$$

Лемма 2 доказана. ■

Пусть f — произвольная функция из P_3, а S — любая схема, реализующая её. Покажем, каким образом по схеме S построить новую схему, которая реализует ту же функцию f, но, возможно (при некоторых условиях на $P(S)$), более надёжно. Для этого возьмём четыре экземпляра схемы S, два элемента $E_{\&}$, один элемент $E_{\vee}$ и построим схему $\psi(S)$, как показано на рис. 2.

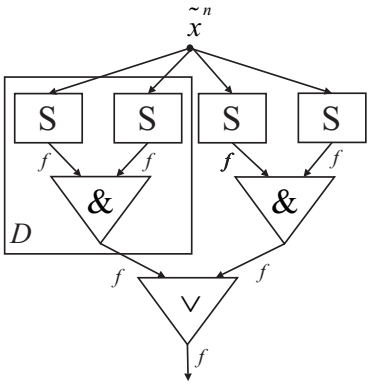

Рис. 2. Схема $\psi(S)$

В теореме 1 найдено рекуррентное соотношение для ненадёжностей схем S и $\psi(S)$.

Теорема 1. Пусть f — произвольная функция из P_3, S — любая схема, реализующая f, а $P(S)$ — ненадёжность схемы S. Тогда схема $\psi(S)$ (рис. 2) реализует функцию f с ненадёжностью

$$P(\psi(S)) \leqslant \varepsilon + (\varepsilon + 2P(S))^2. \tag{1}$$

Доказательство. Пусть f — произвольная функция. Без ограничения общности можно считать, что f зависит от переменных $x_1, \ldots, x_n$. Найдём вероятности ошибок на выходе схемы $\psi(S)$ при всех возможных входных наборах $\tilde{a}^n$ схемы S.

1) Пусть входной набор $\tilde{a}^n$ схемы S такой, что $f(\tilde{a}^n) = 0$. Правильное значение на выходе схемы $\psi(S)$ равно 0. Используя результаты леммы 2, найдём вероятности появления 1 и 2 на выходе схемы $\psi(S)$:

$$P_1(\psi,\tilde{a}^n) \leqslant 2P_1(D,\tilde{a}^n), \quad P_2(\psi,\tilde{a}^n) \leqslant 2P_2(D,\tilde{a}^n).$$

Таким образом,

$$P_{f(\tilde{a}^n)\neq 0}(\psi(S),\tilde{a}^n) = P_1(\psi(S),\tilde{a}^n) + P_2(\psi(S),\tilde{a}^n) \leqslant 2(P_1(D,\tilde{a}^n) + P_2(D,\tilde{a}^n)).$$

Подставив значения для $P_1(D,\tilde{a}^n)$ и $P_2(D,\tilde{a}^n)$ из леммы 1 в случае, когда $f(\tilde{a}^n) = 0$, и учитывая, что $P_1(S,\tilde{a}^n) + P_2(S,\tilde{a}^n) \leqslant P(S)$, получим

$$P_{f(\tilde{a}^n)\neq 0}(\psi(S),\tilde{a}^n) \leqslant 2(P_1^2(S,\tilde{a}^n) + 2P_1(S,\tilde{a}^n)P_2(S,\tilde{a}^n) + P_2^2(S,\tilde{a}^n)) = \\ = 2(P_1(S,\tilde{a}^n) + P_2(S,\tilde{a}^n))^2 \leqslant 2P^2(S).$$

2) Пусть входной набор $\tilde{a}^n$ схемы S такой, что $f(\tilde{a}^n) = 1$. Правильное значение на выходе схемы $\psi(S)$ равно 1. Используя результаты леммы 2, найдём вероятности появления 0 и 2 на выходе схемы $\psi(S)$:

$$P_0(\psi(S),\tilde{a}^n) \leqslant \varepsilon + P_0^2(D,\tilde{a}^n), \quad P_2(\psi(S),\tilde{a}^n) \leqslant 2P_2(D,\tilde{a}^n).$$

Подставив значения для $P_0(D,\tilde{a}^n)$ и $P_2(D,\tilde{a}^n)$ из леммы 1 в случае, когда $f(\tilde{a}^n) = 1$, получим $P_0(\psi(S),\tilde{a}^n) \leqslant \varepsilon + (\varepsilon + 2P_0(S,\tilde{a}^n))^2$, $P_2(\psi(S),\tilde{a}^n) \leqslant 2P_2^2(S,\tilde{a}^n)$. Найдём вероятность ошибки на выходе схемы $\psi(S)$:

$$P_{f(\tilde{a}^n)\neq 1}(\psi(S),\tilde{a}^n) = \varepsilon + (\varepsilon + 2P_0(S,\tilde{a}^n))^2 + 2P_2^2(S,\tilde{a}^n) \leqslant \varepsilon + 2P_2^2(S,\tilde{a}^n) + 4P_0^2(S,\tilde{a}^n) + \\ + 4\varepsilon P_0(S,\tilde{a}^n) + \varepsilon^2 \leqslant \varepsilon + 4P^2(S) + 4\varepsilon P(S) + \varepsilon^2 \leqslant \varepsilon + (\varepsilon + 2P(S))^2$$

(поскольку верно неравенство $P_0(S,\tilde{a}^n) + P_2(S,\tilde{a}^n) \leqslant P(S)$).

3) Пусть входной набор $\tilde{a}^n$ схемы S такой, что $f(\tilde{a}^n) = 2$. Правильное значение на выходе схемы $\psi(S)$ равно 2. Используя результаты леммы 2, найдём вероятности появления 0 и 2 на выходе схемы $\psi(S)$:

$$P_0(\psi(S),\tilde{a}^n) \leqslant \varepsilon + P_0^2(D,\tilde{a}^n), \quad P_1(\psi(S),\tilde{a}^n) \leqslant P_1^2(D,\tilde{a}^n) + 2P_0(D,\tilde{a}^n)P_1(D,\tilde{a}^n).$$

Тогда вероятность ошибки на выходе схемы $\psi(S)$

$$P_{f(\tilde{a}^n)\neq 2}(\psi(S),\tilde{a}^n) = P_0(\psi(S),\tilde{a}^n) + P_1(\psi(S),\tilde{a}^n) \leqslant \varepsilon + [P_0(D,\tilde{a}^n) + P_1(D,\tilde{a}^n)]^2.$$

Подставив значения для $P_0(D,\tilde{a}^n)$ и $P_1(D,\tilde{a}^n)$ из леммы 1 в случае, когда $f(\tilde{a}^n) = 2$, и учитывая, что $P_0(S,\tilde{a}^n) + P_1(S,\tilde{a}^n) \leqslant P(S)$, получим

$$P_{f(\tilde{a}^n)\neq 2}(\psi(S),\tilde{a}^n) \leqslant \varepsilon + (\varepsilon + 2(P_0(S,\tilde{a}^n) + P_1(S,\tilde{a}^n)))^2 \leqslant \varepsilon + (\varepsilon + 2P(S))^2.$$

Поскольку $P(S) = \max\limits_{\tilde{a}^n, f(\tilde{a}^n)=\tau}\{P_{f(\tilde{a}^n)\neq\tau}(\psi(S),\tilde{a}^n)\}$, имеем

$$P(\psi(S)) \leqslant \max\{2P^2(S), \varepsilon + (\varepsilon + 2P(S))^2\} \leqslant \varepsilon + (\varepsilon + 2P(S))^2.$$

Теорема 1 доказана. ■

Используя теорему 1, докажем верхнюю оценку ненадёжности схемы, построенной из элементов, подверженных неисправностям типа 0 на выходах.

3. Верхняя оценка ненадёжности схем

Теорема 2. Любую функцию $f \in P_3$ можно реализовать такой схемой S, что $P(S) \leqslant 3\varepsilon$ при всех $\varepsilon \in (0, 1/300]$.

Доказательство. Индукция по n — числу переменных функции $f(\tilde{x}^n)$.

1. Докажем утверждение для $n = 1$, т. е. для всех возможных функций $f(x)$, зависящих от одной переменной x. Разложим функцию $f(x)$ по переменной x [19, с. 47]:

$$f(x) = J_0(x)\&f(0) \vee J_1(x)\&f(1) \vee J_2(x)\&f(2).$$

Чтобы промоделировать эту формулу схемой (обозначим её S'), достаточно 11 элементов. Поэтому ненадёжность $P(S')$ схемы S' удовлетворяет неравенству $P(S') \leqslant 11\varepsilon$.

По схеме S' построим схему $\psi(S')$, как показано на рис. 2. Используя соотношение (1) из теоремы 1 и условие $\varepsilon \leqslant 1/300$, оценим ненадёжность схемы $\psi(S')$:

$$P(\psi(S')) \leqslant \varepsilon + (2 \cdot 11\varepsilon + \varepsilon)^2 = \varepsilon + 529\varepsilon^2 \leqslant \varepsilon + \frac{529}{300}\varepsilon \leqslant 3\varepsilon.$$

Следовательно, $\psi(S')$ — искомая схема S. Для $n = 1$ теорема верна.

2. Пусть утверждение верно для функций $f(\tilde{x}^{n-1})$ с числом переменных $n - 1$. Докажем, что оно верно для функций $f(\tilde{x}^n)$. Разложим функцию $f(x_1, \ldots, x_{n-1}, x_n)$ по переменной x_n:

$$f(x_1, \ldots, x_{n-1}, x_n) = J_0(x_n)\&f(x_1, \ldots, x_{n-1}, 0)\vee$$
$$\vee J_1(x_n)\&f(x_1, \ldots, x_{n-1}, 1) \vee J_2(x_n)\&f(x_1, \ldots, x_{n-1}, 2).$$

Используя это разложение, построим схему C (рис. 3), реализующую функцию $f(\tilde{x}^n)$, причём S_0 — схема, реализующая функцию $f_0 = f(x_1, \ldots, x_{n-1}, 0)$, S_1 — схема, реализующая функцию $f_1 = f(x_1, \ldots, x_{n-1}, 1)$, а S_2 — схема, реализующая функцию $f_2 = f(x_1, \ldots, x_{n-1}, 2)$.

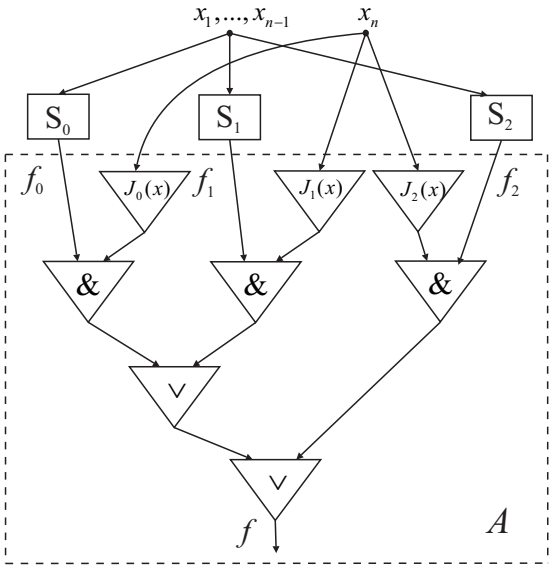

Рис. 3. Схема C

В схеме C выделим подсхему A, состоящую из восьми элементов, выход которой является выходом схемы C, а на входы подаются значения x_n, $f_0 = f(x_1, \ldots, x_{n-1}, 0)$, $f_1 = f(x_1, \ldots, x_{n-1}, 1)$ и $f_2 = f(x_1, \ldots, x_{n-1}, 2)$.

Подсхема A состоит из восьми элементов, поэтому её ненадёжность $P(S) \leqslant 8\varepsilon$. Функции $f_0 = f(x_1, \ldots, x_{n-1}, 0)$, $f_1 = f(x_1, \ldots, x_{n-1}, 1)$ и $f_2 = f(x_1, \ldots, x_{n-1}, 2)$ по индуктивному предположению можно реализовать такими схемами S_0, S_1 и S_2, что ненадёжность каждой из них не больше 3ε. Если схема A исправна, то для реализации f она использует значение одной из схем, реализующих функции f_0, f_1 и f_2. Поэтому $P(C) \leqslant P(A) + 3\varepsilon \leqslant 8\varepsilon + 3\varepsilon = 11\varepsilon$.

По схеме C построим схему $\psi(C)$ (см. рис. 2). Воспользуемся соотношением (1) и оценим ненадёжность схемы $\psi(C)$, учитывая, что $\varepsilon \leqslant 1/300$:

$$P(\psi(C)) \leqslant \varepsilon + (2 \cdot 11\varepsilon + \varepsilon)^2 = \varepsilon + 529\varepsilon^2 \leqslant \varepsilon + \frac{529}{300}\varepsilon \leqslant 3\varepsilon.$$

Следовательно, схема $\psi(C)$ — искомая схема S. Теорема 2 доказана. ■

Теорема 3. Любую функцию $f \in P_3$ можно реализовать такой схемой S, что $P(S) \leqslant \varepsilon + 12\varepsilon^2$ при всех $\varepsilon \in (0, 1/300]$.

Доказательство. По теореме 2 любую функцию f можно реализовать схемой D с ненадёжностью $P(D) \leqslant 3\varepsilon$. По схеме D построим схему $\psi(D)$ (см. рис. 2) и оценим её ненадёжность по формуле (1) из теоремы 1:

$$P(\psi(D)) \leqslant \varepsilon + (2 \cdot 3\varepsilon + \varepsilon)^2 = \varepsilon + 49\varepsilon^2 \leqslant 1,17\varepsilon$$

при $\varepsilon \in (0, 1/300]$. По схеме $\psi(D)$ построим схему $\psi^2(D)$ и оценим её ненадёжность по формуле (1) из теоремы 1: $P(\psi^2(D)) \leqslant \varepsilon + (\varepsilon + 2,34\varepsilon)^2 \leqslant \varepsilon + 12\varepsilon^2$. Схема $\psi^2(D)$ — искомая схема S. ■

Из теоремы 3 следует, что любую функцию из P_3 можно реализовать схемой, ненадёжность которой асимптотически (при $\varepsilon \to 0$) не больше ε.

4. Нижняя оценка ненадёжности схем

Пусть $K(n)$ — множество функций трёхзначной логики, каждая из которых зависит от переменных $x_1, \ldots, x_n$ $(n \geqslant 1)$, отлична от константы 0 и функций $x_1, \ldots, x_n$. Обозначим $K = \bigcup\limits_{n=1}^{\infty} K(n)$. Очевидно, что $|K(n)| = 3^{3^n} - n - 1$, а значит, класс $K(n)$ содержит почти все функции из множества $P_3(n)$ $\left(\text{поскольку } \lim\limits_{n\to\infty} \dfrac{3^{3^n} - n - 1}{3^{3^n}} = 1\right)$.

Справедлива теорема 4 о нижней оценке ненадёжности схем, каждая из которых реализует функцию из класса K.

Теорема 4. Пусть функция $f \in K$. Тогда для любой схемы S, реализующей f, верно неравенство $P(S) \geqslant \varepsilon$.

Доказательство. Без ограничения общности можно считать, что функция $f \in K(n)$, пусть S — любая схема, реализующая f. Заметим, что схема S содержит хотя бы один элемент. Пусть E — функциональный элемент схемы S, выход которого является выходом схемы.

Поскольку $f \not\equiv 0$, найдутся такие значение $i \in E_3 \setminus \{0\}$ и входной набор $\tilde{a}^n$, что $f(\tilde{a}^n) = i$. Вычислим вероятность $P_0(S, \tilde{a}^n)$ появления значения 0 на выходе схемы S на наборе $\tilde{a}^n$, обозначив через p_0 вероятность появления нулевого набора на входах элемента E: $P_0(S, \tilde{a}^n) = p_0 + (1 - p_0)\varepsilon = \varepsilon + p_0(1 - \varepsilon) \geqslant \varepsilon$. ■

Из теоремы 4 следует, что ненадёжность любой схемы, реализующей функцию $f \in K$, не меньше ε. Это означает, что схема, реализующая функцию $f \in K$ и удовле-

творяющая условиям теоремы 3, является асимптотически оптимальной по надёжности и функционирует с ненадёжностью, асимптотически равной ε при $\varepsilon \to 0$.

Напомним, что функции 0, x_i ($i \in \mathbb{N}$) (и именно эти функции не содержатся в классе K) можно реализовать абсолютно надёжно.

Заключение

В базисе Россера — Туркетта (в P_3) при неисправностях типа 0 на выходах элементов:

1) любую функцию трёхзначной логики можно реализовать схемой, ненадёжность которой асимптотически (при $\varepsilon \to 0$) не больше ε;

2) для любой функции $f \in K$ такая схема является асимптотически оптимальной по надёжности и функционирует с ненадёжностью, асимптотически равной ε при $\varepsilon \to 0$;

3) функции $f \notin K$ (т. е. функции 0, x_i ($i \in \mathbb{N}$)) можно реализовать абсолютно надёжно.

ЛИТЕРАТУРА

1. *Von Neuman J.* Probabilistic logics and the synthesis of reliable organisms from unreliable components // Automata Studies / eds. C. Shannon and J. McCarthy. Princeton University Press, 1956. P. 43–98. (Русс. пер.: Автоматы. М.: ИЛ, 1956. С. 68–139.)
2. *Добрушин Р. Л., Ортюков С. И.* Верхняя оценка для избыточности самокорректирующихся схем из ненадежных функциональных элементов // Пробл. передачи информ. 1977. Т. 13. № 3. С. 56–76.
3. *Ортюков С. И.* Об избыточности реализации булевых функций схемами из ненадежных элементов // Труды семинара по дискретной математике и её приложениям (Москва, 27–29 января 1987 г.). М.: Изд-во МГУ, 1989. С. 166–168.
4. *Uhlig D.* Reliable networks from unreliable gates with almost minimal complexity // LNCS. 1987. V. 278. P. 462–469.
5. *Pippenger N.* On networks of noisy gates // 26th Ann. Symp. Foundations of Computer Science. Portland, 21–23 Oct. 1985. P. 30–38.
6. *Яблонский С. В.* Асимптотически наилучший метод синтеза надежных схем из ненадежных элементов // Banach Center Publ. 1982. V. 7. No. 1. P. 11–19.
7. *Тарасов В. В.* К синтезу надежных схем из ненадежных элементов // Матем. заметки. 1976. Т. 20. № 3. С. 391–400.
8. *Виноградов Ю. А.* О синтезе трехзначных МДП-схем // Математические вопросы кибернетики. 1991. Вып. 3. С. 187–198.
9. *Виноградов Ю. А.* О синтезе четырехзначных квазикомплементарных МОП-схем // Математические вопросы кибернетики. 1999. Вып. 8. С. 298–300.
10. *Барсукова О. Ю.* Синтез надежных схем, реализующих функции двузначной и трехзначной логик: дис. ... канд. физ.-мат. наук. Пенза, 2014. 87 с.
11. *Алехина М. А., Каргин С. П.* О синтезе схем из ненадежных элементов в P_4 // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. 2014. № 4 (32). С. 47–56.
12. *Алехина М. А.* Синтез схем из ненадежных элементов в P_k // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. 2015. № 3. С. 8–10.
13. *Алехина М. А., Барсукова О. Ю.* О надежности схем, реализующих функции из P_3 // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. 2012. № 1 (21). С. 57–65.

14. *Алехина М. А., Барсукова О. Ю.* Оценки ненадежности схем в базисе Россера — Туркетта // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. 2014. № 1 (29). С. 5–19.
15. *Алехина М. А., Барсукова О. Ю.* Ненадёжность схем в базисе Россера — Туркетта // Прикладная дискретная математика. Приложение. 2014. № 7. С. 109–110.
16. *Алехина М. А., Барсукова О. Ю.* О надежности схем, реализующих функции трехзначной логики // Дискрет. анализ и исслед. опер. 2014. Т. 21. № 4 (118). С. 12–24.
17. *Алехина М. А., Каргин С. П.* Асимптотически оптимальные по надежности схемы в базисе Россера — Туркетта в P_4 // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. 2015. № 1. С. 37–53.
18. *Алехина М. А., Барсукова О. Ю.* Нижняя оценка ненадёжности схем в базисе, состоящем из функции Вебба // Прикладная дискретная математика. Приложение. 2015. № 8. С. 102–103.
19. *Яблонский С. В.* Введение в дискретную математику. М.: Высш. шк., 2001. 384 с.

REFERENCES

1. *Von Neuman J.* Probabilistic logics and the synthesis of reliable organisms from unreliable components. Automata Studies. Eds. C. Shannon and J. McCarthy. Princeton University Press, 1956, pp. 43–98.
2. *Dobrushin R. L. and Ortyukov S. I.* Upper bound on the redundancy of self-correcting arrangements of unreliable functional elements. Problems Inform. Transmission, 1977, vol. 13, iss. 3, pp. 203–218.
3. *Ortyukov S. I.* Ob izbytochnosti realizatsii bulevykh funktsiy skhemami iz nenadezhnykh elementov [On the redundancy of the Boolean functions implementation by circuits from unreliable elements]. Proc. Seminar on Discr. Math. and its Appl. (Moscow, 27–29 Jan. 1987). Moscow, MSU Publ., 1989, pp. 166–168. (in Russian)
4. *Uhlig D.* Reliable networks from unreliable gates with almost minimal complexity. LNCS, 1987, vol. 278, pp. 462–469.
5. *Pippenger N.* On networks of noisy gates. 26th Ann. Symp. Foundations of Computer Science, Portland, 21–23 Oct. 1985, pp. 30–38.
6. *Yablonskiy C. V.* Asimptoticheski nailuchshiy metod sinteza nadezhnykh skhem iz nenadezhnykh elementov [Asymptotically best method for synthesizing reliable circuits from unreliable elements]. Banach Center Publ., 1982, vol. 7, no. 1, pp. 11–19. (in Russian)
7. *Tarasov V. V.* The synthesis of reliable circuits from unreliable elements. Math. Notes, 1976, vol. 20, iss. 3, pp. 775–780.
8. *Vinogradov Yu. A.* O sinteze trekhznachnykh MDP-skhem [On the synthesis of three-valued MDS-circuits]. Matematicheskie Voprosy Kibernetiki, 1991, iss. 3, pp. 187–198. (in Russian)
9. *Vinogradov Yu. A.* O sinteze chetyrekhznachnykh kvazikomplementarnykh MOP-skhem [On the synthesis of four-valued quasicomplementary MOSFETs]. Matematicheskie Voprosy Kibernetiki, 1999, iss. 8, pp. 298–300. (in Russian)
10. *Barsukova O. Yu.* Sintez nadezhnykh skhem, realizuyushchikh funktsii dvuznachnoy i trekhznachnoy logik [Synthesis of reliable schemes realizing the functions of two-valued and three-valued logics]. PhD Thesis, Penza, 2014. 87 p. (in Russian)
11. *Alekhina M. A. and Kargin S. P.* O sinteze skhem iz nenadezhnykh elementov v P_4 [On synthesis of unreliable element circuits in P_4]. Izvestiya Vysshikh Uchebnykh Zavedeniy. Povolzhskiy Region. Fiziko-Matematicheskie Nauki, 2014, no. 4 (32), pp. 47–56. (in Russian)
12. *Alekhina M. A.* Sintez skhem iz nenadezhnykh elementov v P_k [Synthesis of circuits containing unreliable gates in P_k]. Izvestiya Vysshikh Uchebnykh Zavedeniy. Povolzhskiy Region. Fiziko-Matematicheskie Nauki, 2015, no. 3, pp. 8–10. (in Russian)

13. *Alekhina M. A. and Barsukova O. Yu.* O nadezhnosti skhem, realizuyushchikh funktsii iz P_3 [On the reliability of circuits realizing functions from P_3]. Izvestiya Vysshikh Uchebnykh Zavedeniy. Povolzhskiy Region. Fiziko-Matematicheskie Nauki, 2012, no. 1 (21), pp. 57–65. (in Russian)
14. *Alekhina M. A. and Barsukova O. Yu.* Otsenki nenadezhnosti skhem v bazise Rossera — Turketta [Circuit failure estimate in the Rosser — Turkett basis]. Izvestiya Vysshikh Uchebnykh Zavedeniy. Povolzhskiy Region. Fiziko-Matematicheskie Nauki, 2014, no. 1 (29), pp. 5–19. (in Russian)
15. *Alekhina M. A. and Barsukova O. Yu.* Nenadezhnost' skhem v bazise Rossera — Turketta [Unreliability of circuits in the basis by Rosser — Turkett]. Prikladnaya Diskretnaya Matematika. Prilozhenie, 2014, no. 7, pp. 109–110. (in Russian)
16. *Alekhina M. A. and Barsukova O. Yu.* O nadezhnosti skhem, realizuyushchikh funktsii trekhznachnoy logiki [On reliability of circuits realizing ternary logic functions]. Diskretn. Anal. Issled. Oper., 2014, vol. 21, no. 4, pp. 12–24. (in Russian)
17. *Alekhina M. A. and Kargin S. P.* Asimptoticheski optimal'nye po nadezhnosti skhemy v bazise Rossera — Turketta v P_4 [Asymptotic reliability-optimal circuits in the Rosser — Turkett basis in P_4]. Izvestiya Vysshikh Uchebnykh Zavedeniy. Povolzhskiy Region. Fiziko-Matematicheskie Nauki, 2015, no. 1, pp. 37–53. (in Russian)
18. *Alekhina M. A. and Barsukova O. Yu.* Nizhnyaya otsenka nenadezhnosti skhem v bazise, sostoyashchem iz funktsii Vebba [A lower bound for unreliability of circuits in the Webb basis]. Prikladnaya Diskretnaya Matematika. Prilozhenie, 2015, no. 8, pp. 102–103. (in Russian)
19. *Yablonskiy S. V.* Vvedenie v diskretnuyu matematiku [Introduction to Discrete Mathematics]. Moscow, Vyssh. Shk., 2001. 384 p. (in Russian)

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ И ПРОГРАММИРОВАНИЯ

УДК 519.682

О ПРИМЕНЕНИИ МНОГОМЕРНОГО КОМПЛЕКСНОГО АНАЛИЗА В ТЕОРИИ ФОРМАЛЬНЫХ ЯЗЫКОВ И ГРАММАТИК[1]

О. И. Егорушкин, И. В. Колбасина, К. В. Сафонов

Сибирский государственный университет науки и технологий им. акад. М. Ф. Решетнёва, г. Красноярск, Россия

Исследуются системы полиномиальных уравнений над полукольцом (относительно символов с некоммутативным умножением и коммутативным сложением). Такие системы уравнений интерпретируются как грамматики формальных языков и решаются относительно нетерминальных символов в виде формальных степенных рядов, зависящих от терминальных символов. Рассматривается коммутативный образ системы уравнений в предположении, что символы являются переменными, принимающими значения из поля комплексных чисел. Устанавливаются связи между решениями системы некоммутативных символьных уравнений и её коммутативного образа, тем самым методы многомерного комплексного анализа привлекаются в теорию формальных языков и грамматик. Доказывается дискретный аналог теоремы о неявном отображении для формальных грамматик: достаточным условием существования и единственности решения системы некоммутативных уравнений в виде формальных степенных рядов является отличие от нуля якобиана коммутативного образа этой системы. Предложен также новый метод синтаксического анализа мономов контекстно-свободного языка как модели языков программирования, основанный на интегральном представлении синтаксического многочлена программы. При этом показано, что интеграл фиксированной кратности по циклу позволяет найти синтаксический многочлен монома (программы) с неограниченным числом символов, что даёт новый подход к проблеме синтаксического анализа.

Ключевые слова: *формальный степенной ряд, коммутативный образ, синтаксический анализ, интегральное представление.*

DOI 10.17223/20710410/37/6

[1]Исследование выполнено при финансовой поддержке РФФИ и Правительства Красноярского края в рамках научного проекта № 17-47-240318.

ON APPLICATION OF MULTIDIMENSIONAL COMPLEX ANALYSIS IN FORMAL LANGUAGE AND GRAMMAR THEORY

O. I. Egorushkin, I. V. Kolbasina, K. V. Safonov

Reshetnev State University of Science and Technology, Krasnoyarsk, Russia

E-mail: safonovkv@rambler.ru

Systems of polynomial equations over a semiring (with respect to symbols with a non-commutative multiplication and a commutative addition) are investigated. These systems of equations are interpreted as the grammars of formal languages and are resolved with respect to the non-terminal symbols in the form of the formal power series (FPS) depending on the terminal symbols. The commutative image of the system of equations is determined under the assumption that the symbols are variables taking values from the field of complex numbers. The connections between solutions of the non-commutative symbolic system of equations and its commutative image are established, thus the methods for multidimensional complex analysis are involved in the theory of formal language grammars. A discrete analogue of implicit mapping theorem onto formal grammars is proved: a sufficient condition for the existence and the uniqueness of a solution of a non-commutative system of equations in the form of FPS is the inequality to zero of the Jacobian of the commutative image of this system. A new method for syntactic analysis of the monomials of a context-free language as a model of programming languages is also proposed. The method is based on the integral representation of the syntactic polynomial of a program. It is shown that the integral of a fixed multiplicity over a cycle allows finding the syntactic polynomial of a monomial (program) with the unlimited number of symbols, that gives a new approach to the problem of syntactic analysis.

Keywords: *formal power series, commutative image, syntactic analysis, integral representation.*

Введение

В теории формальных языков и грамматик исходным объектом является алфавит $\{z_1, \ldots, z_n, x_1, \ldots, x_m\}$, над которым определена некоммутативная операция умножения (конкатенации) и коммутативная операция формальной суммы; алфавит вместе с операциями образует полукольцо. Символы $x_1, \ldots, x_m$ называются терминальными символами и составляют словарь формального языка, а символы $z_1, \ldots, z_n$ — нетерминальными и нужны для задания совокупности грамматических правил (грамматики), которые порождают язык. По этим правилам определяются «правильные» мономы от терминальных символов $x_1, \ldots, x_m$, которые являются правильными предложениями языка. Кроме того, определена коммутативная операция умножения символов алфавита на комплексные числа, что позволяет рассматривать символьные многочлены и формальные степенные ряды (ФСР) с числовыми коэффициентами. При этом формальным языком является такой ФСР, членами которого являются все правильные мономы, определённые данной грамматикой [1, 2].

Заметим, что порождающую грамматику для важных классов формальных языков можно записать в виде системы символьных полиномиальных уравнений

$$P_j(z, x) = 0,\ P_j(0, 0) = 0,\ j = 1, \ldots, k, \tag{1}$$

которая решается относительно символов $(z_1, \ldots, z_n) = z$ в виде ФСР, зависящих от символов $(x_1, \ldots, x_m) = x$: $z = z(x) = (z_1(x), \ldots, z_n(x))$; при этом ФСР $z_1(x)$ является формальным языком, который определяется этой грамматикой.

Так, для класса контекстно-свободных языков (кс-языков) система символьных уравнений (1) может быть представлена в виде системы уравнений Хомского — Щутценберже

$$z_j - Q_j(z, x) = 0,\ j = 1, \ldots, n, \tag{2}$$

где правые части удовлетворяют естественным условиям: $Q_j(0, 0) = 0$, многочлены $Q_j(z, 0)$ не содержат линейных членов [1]. Отметим, что классу кс-языков принадлежит большинство языков программирования [2].

Более широкий класс языков непосредственно составляющих задаётся системой символьных уравнений

$$m_j(z, x) - M_j(z, x) = 0,\ j = 1, \ldots, k,$$

где $m_j(z, x)$ — моном, а $M_j(z, x)$ — многочлен [2], которая также является частным случаем системы символьных уравнений (1).

Наконец, отметим языки в нормальной форме Грейбах, которые также можно задать системой символьных полиномиальных уравнений (1), где все многочлены $P_j(z, x)$ имеют степень не выше второй, вводя при необходимости новые терминальные и нетерминальные символы [2].

Таким образом, исследуя систему символьных полиномиальных уравнений (1), можно изучать свойства порождаемых ею формальных *полиномиальных* языков, в том числе языков из наиболее важных классов. Однако свойства некоммутативной системы уравнений (1) изучены мало. Так, в частном случае системы уравнений Хомского — Щутценберже (2) известно, что она имеет единственное решение $z = (z_1(x), \ldots, z_n(x))$ в виде ФСР, но в общем случае условия разрешимости системы уравнений (1) неизвестны.

Для исследования нет подходящих методов: исключению неизвестных препятствует некоммутативность умножения и отсутствие операции деления. Например, методы дискретной математики, в частности теории графов, традиционно играющие важную роль в теории формальных языков и грамматик, не вполне приспособлены для исследования некоммутативной системы уравнений (1).

В случае коммутативных неизвестных системы полиномиальных уравнений изучаются в рамках алгебраической геометрии методами коммутативной алгебры. Если при этом рассматривать неизвестные над алгебраически замкнутым полем комплексных чисел, то можно использовать также методы многомерного комплексного анализа [3]. Однако методы коммутативной алгебры и комплексного анализа не выглядят удобными для изучения систем полиномиальных уравнений (1) с неизвестными, некоммутативными по умножению.

Таким образом, дальнейшее развитие теории формальных полиномиальных языков и грамматик (каждая грамматика может быть рассмотрена как система таких уравнений) затрудняется отсутствием подходящих методов, возникающих как в дискретной математике, так и в коммутативной алгебре и комплексном анализе. И всё же привлечь в теорию формальных языков и грамматик методы многомерного комплексного анализа можно на основе следующего подхода, предложенного в работах [4, 5].

Этот подход состоит в том, что наряду с некоммутативной системой (1) рассматривается её коммутативный образ, который получается в предположении, что все

переменные, входящие в систему, принимают значения из поля комплексных чисел. Коммутативный образ этой системы уравнений является системой полиномиальных уравнений в пространстве $\mathbb{C}^{n+m}_{z,x}$, а коммутативный образ формального языка становится кратным степенным рядом, представляющим алгебраическую функцию в $\mathbb{C}^m_x$, которая является решением этой системы. Далее можно исследовать коммутативный образ системы уравнений (1), применяя методы многомерного комплексного анализа. Наконец, остаётся установить, какие свойства некоммутативной системы уравнений (1) вытекают из свойств её коммутативного образа [4, 5].

В рамках этого подхода было отмечено, что из совместности некоммутативной системы (1) следует совместность её коммутативного образа, однако обратное утверждение неверно [4, 5]. В результате вопрос об исследовании некоммутативной системы (1) по свойствам её коммутативного образа оставался открытым.

Цель данной работы состоит в том, чтобы сделать следующий шаг по реализации указанного подхода, применяя в теории формальных языков и грамматик новые для неё методы многомерного комплексного анализа: во-первых, получить достаточное условие совместности (и единственности решения в виде ФСР) системы уравнений (1); во-вторых, решить важную прикладную задачу синтаксического анализа любой программы, написанной на кс-языке программирования.

Ниже доказано достаточное условие разрешимости некоммутативной системы (1) в терминах якобиана её коммутативного образа, а также предложен новый подход к проблеме синтаксического анализа, состоящий в том, что можно осуществить синтаксический анализ программы, переходя от «программирования к интегрированию» — вычислив кратный интеграл по циклу от некоторой рациональной формы.

1. Коммутативный образ ФСР

Упорядочим члены ФСР, следуя [4, 5]. Пусть все мономы от $x_1, \ldots, x_m$ сгруппированы в однородные многочлены, расположенные по возрастанию степеней, затем перенумеруем мономы каждого из многочленов в лексикографическом порядке, переходя от меньшей степени к большей. При таком упорядочивании все мономы от символов $x_1, \ldots, x_m$ единственным образом записываются в виде последовательности $u_0, u_1, \ldots$, играющей роль базиса ФСР. Теперь каждый ряд s можно однозначно записать в виде разложения по этому базису:

$$s = \sum_{i=0}^{\infty} \langle s, u_i \rangle u_i, \tag{3}$$

где $\langle s, u_i \rangle$ — числовой коэффициент при мономе u_i.

Поставим в соответствие ФСР (3) его коммутативный образ $\mathrm{ci}(s)$ — степенной ряд, который получается из s в предположении, что символы $x_1, \ldots, x_m$ (равно как и $z_1, \ldots, z_n$) обозначают коммутативные переменные, принимающие значения из поля комплексных чисел.

В этом предположении любой моном u_i от символов $x_1, \ldots, x_m$ можно записать в виде $x_1^{\alpha_1} \cdot \ldots \cdot x_m^{\alpha_m}$, где $\alpha_j = \deg_{x_j}(u_i)$ — число вхождений (степень) символа x_j в этот моном. Если обозначить мультииндекс $\alpha = (\alpha_1, \ldots, \alpha_m)$, то можно записать равенство $\alpha = \deg_x(u_i)$, с учётом которого получаются следующие равенства:

$$\mathrm{ci}(s) = \sum_{i=0}^{\infty} \langle s, u_i \rangle \mathrm{ci}(u_i) = \sum_{\alpha} \left(\sum_{\alpha = \deg_x(u_i)} \langle s, u_i \rangle \right) x^{\alpha} \stackrel{\text{def}}{=} \sum_{\alpha} c_{\alpha} x^{\alpha}.$$

Впервые коммутативный образ ФСР рассмотрел А. Л. Семёнов [6], используя методы вещественного анализа для решения алгоритмических проблем, связанных с кс-

языками. Эта работа открыла дорогу для приложений как вещественного, так и многомерного комплексного анализа в теории формальных языков и грамматик.

Например, коммутативный образ ФСР может быть использован при решении проблемы академика В. М. Глушкова: установить критерии, с помощью которых можно выяснить, является данный ФСР кс-языком или нет [2, с. 196]. В самом деле, коммутативный образ кс-языка является кратным степенным рядом, представляющим алгебраическую функцию многих комплексных переменных, поэтому можно получить частичное решение этой проблемы, используя критерии алгебраичности для суммы степенного ряда [7 – 10].

Далее рассмотрим коммутативный образ системы символьных полиномиальных уравнений, представляющий собой систему полиномиальных уравнений в $\mathbb{C}^{n+m}_{z,x}$.

Теорема 1. Если $z = z(x) = (z_1(x), \ldots, z_n(x))$ — решение некоммутативной системы уравнений (1) в виде символьных ФСР, то коммутативные ФСР $z = \mathrm{ci}(z(x)) = (\mathrm{ci}(z_1(x)), \ldots, \mathrm{ci}(z_n(x)))$ над полем комплексных чисел сходятся в окрестности нуля, определяя ростки голоморфных алгебраических функций, и являются решением коммутативной системы уравнений

$$\mathrm{ci}(P_j(z,x)) = 0,\ \mathrm{ci}(P_j(0,0)) = 0,\ j = 1, \ldots, k. \tag{4}$$

Доказательство. Пусть решение $z(x) = (z_1(x), \ldots, z_n(x))$ системы (1), представленное ФСР, имеет вид

$$z_1(x) = \sum_i \langle z_1, u_i\rangle u_i,\ \ldots,\ z_n(x) = \sum_i \langle z_n, u_i\rangle u_i.$$

Отметим, что, подставляя это решение в многочлен $P_j(z,x)$, стоящий в левой части уравнений (1), получаем ФСР в виде разложения по универсальному базису с нулевыми коэффициентами:

$$l_j = \sum_i \langle l_j, u_i\rangle u_i = \sum_i 0 \cdot u_i,\ j = 1, \ldots, k.$$

Тогда имеют место равенства

$$\mathrm{ci}(P_j(z,x))|_{z=\mathrm{ci}(z(x))} = \mathrm{ci}(P_j(z(x),x)) = \mathrm{ci}(l_j) = \sum_i 0 \cdot \mathrm{ci}(u_i),\ j = 1, \ldots, k,$$

т. е. совокупность коммутативных ФСР удовлетворяет системе уравнений (4).

Покажем, что все ФСР $\mathrm{ci}(z(x))$ сходятся в некоторой окрестности нуля. В самом деле, если ФСР над полем комплексных чисел удовлетворяет уравнению с голоморфными коэффициентами (функциями, представленными абсолютно сходящимися в окрестности нуля степенными рядами), то он также абсолютно сходится в некоторой окрестности нуля, представляя голоморфную функцию [11]. Следовательно, коммутативные ФСР $\mathrm{ci}(z(x))$ сходятся в некоторой окрестности нуля, порождая ростки алгебраической вектор-функции, а решения (z,x) системы (4), рассматриваемые как точки в комплексном пространстве $\mathbb{C}^{n+m}_{z,x}$, можно записать с помощью этой вектор-функции в виде $(z(x),x)$. ■

Частный случай теоремы 1, когда $k = n$, рассмотрен в [5].

Подчеркнём, что совместность понимается для некоммутативной системы (1) и коммутативной системы (4) по-разному: в первом случае решением являются символьные ФСР от $x_1, \ldots, x_m$, а во втором — точки в комплексном пространстве, параметризованные алгебраическим отображением $z = z(x)$; его голоморфные в нуле ветви

являются коммутативными образами символьных ФСР, представляющих компоненты решения системы (1). По теореме 1, если некоммутативная система (1) совместна, то коммутативная система (4) имеет голоморфное в начале координат решение. Обратное, вообще говоря, неверно. В самом деле, система уравнений

$$z_1 - z_2 = x_1x_2,\ z_1 - z_2 = x_2x_1$$

является несовместной, тем не менее её коммутативный образ имеет решение: $z_1 = s + x_1x_2$, $z_2 = s$, где s — произвольный коммутативный ФСР. Таким образом, можно сделать следующие замечания.

Замечание 1. Множество решений системы уравнений (4), вообще говоря, шире, чем множество коммутативных образов решений системы уравнений (1), и из совместности системы уравнений (4) не следует совместность исходной некоммутативной системы (1).

Замечание 2. Естественная гипотеза о том, что система (1) совместна тогда и только тогда, когда система (4) имеет голоморфное в начале координат решение $z = z(x)$, неверна.

Рангом матрицы A называется наибольшее число её строк, линейно независимых над полем комплексных чисел; обозначается $\mathrm{rank}(A)$.

Теорема 2. Пусть $A = (a_{ij}(x))_{k\times n}$ — матрица, элементы которой являются многочленами от символьных некоммутативных переменных $x = (x_1, \ldots, x_m)$, и матрица $\mathrm{ci}(A) = (\mathrm{ci}(a_{ij}(x)))_{k\times n}$ является коммутативным образом матрицы A. Тогда выполнено неравенство

$$\mathrm{rank}(\mathrm{ci}(A)) \leqslant \mathrm{rank}(A).$$

Доказательство. Пусть, для определённости, линейно независимыми являются строки с номерами $1, \ldots, r$, следовательно, $\mathrm{rank}(A) = r$. Приравнивая к нулю линейную комбинацию этих строк с коэффициентами $z_1, \ldots, z_r$, получим систему полиномиальных уравнений

$$z_1a_{1j}(x) + \ldots + z_ra_{rj}(x) = 0,\ j = 1, \ldots, n,$$

которая имеет единственное решение в виде ФСР: $z_1 = \ldots = z_r = 0$. Коммутативный образ этой системы уравнений

$$z_1\mathrm{ci}(a_{1j}(x)) + \ldots + z_r\mathrm{ci}(a_{rj}(x)) = 0,\ j = 1, \ldots, n,$$

в соответствии с замечанием 1, может иметь более широкое множество решений, включающее решение $z_1 = \ldots = z_r = 0$. В этом случае коммутативные образы исходных строк являются линейно зависимыми, следовательно, $\mathrm{rank}(\mathrm{ci}(A)) \leqslant \mathrm{rank}(A)$. ■

Отметим, что теореме 1 эквивалентна следующая теорема.

Теорема 3. Если коммутативная система уравнений (4) не имеет решения, голоморфного в начале координат, то некоммутативная система (1) несовместна.

Таким образом, условия несовместности системы уравнений (4) также представляют интерес в теории формальных языков и грамматик.

2. Дискретный аналог теоремы о неявном отображении

Конечно, наибольший интерес для приложений представляют условия, которые обеспечивают совместность системы некоммутативных символьных уравнений (1), а также единственность её решения. Согласно замечанию 2, таким условием не может

быть существование голоморфного в начале координат решения системы коммутативных уравнений (4). Тем не менее это условие можно получить с помощью такого инструмента, как якобиан системы функций.

Рассмотрим систему уравнений (1) в случае, когда $k = n$. Пусть

$$J(z, x) = \det((\mathrm{ci}(P_i(z, x)))'_{z_j})$$

— якобиан системы уравнений (4) относительно переменных $z_1, \ldots, z_n$.

Дискретным (символьным) аналогом теоремы о неявном отображении является следующая теорема.

Теорема 4. Если для некоммутативной символьной системы уравнений (1) выполнено неравенство $J(0, 0) \neq 0$, то она имеет единственное решение в виде ФСР.

Замечание 3. Неравенство является условием теоремы о неявном отображении для коммутативной системы уравнений (4) с переменными в $\mathbb{C}^{n+m}_{z,x}$ и влечёт существование и единственность её голоморфного решения; тем не менее оказывается, что это неравенство влечёт также существование и единственность решения исходной некоммутативной символьной системы уравнений (1).

Доказательство. Обозначим $L_1(z), \ldots, L_n(z)$ линейные части многочленов $P_1(z, x), \ldots, P_n(z, x)$ соответственно, зависящие только от $z_1, \ldots, z_n$:

$$L_j(z) = a_{j1}z_1 + \ldots + a_{jn}z_n, \ j = 1, \ldots, n.$$

Далее, обозначая многочлены $S_j(z, x) = L_j(z) - P_j(z, x)$, запишем исходную систему уравнений в виде

$$a_{j1}z_1 + \ldots + a_{jn}z_n = S_j(z, x), \ j = 1, \ldots, n. \tag{5}$$

Рассмотрим число $J(0, 0)$; легко видеть, что имеют место равенства

$$J(0, 0) = \det((\mathrm{ci}(L_i(z)))'_{z_j}) = \det(((L_i(z))'_{z_j}) = \det(a_{ij}).$$

Поскольку $\det(a_{ij}) \neq 0$, матрицу (a_{ij}) можно привести к диагональному виду, умножая на числа уравнения системы (5) и складывая их [12]. При этом система (5) перейдёт в эквивалентную систему уравнений Хомского — Щутценберже (2), в которой правые части уравнений не содержат линейных членов.

Как отмечалось выше, эта система имеет единственное решение, которое можно получить методом последовательных приближений [1]:

$$z^{(k+1)}(x) = Q(z^{(k)}(x), x), \ k = 0, 1, \ldots; \ z^{(0)} = 0;$$
$$Q(z, x) = (Q_1(z, x), \ldots, Q_n(z, x)); \ z(x) = \lim_{k\to\infty} z^{(k)}(x).$$

Теорема доказана. ■

Замечание 4. Неравенство $J(0, 0) \neq 0$ обусловлено свойствами линейных многочленов $L_1(z), \ldots, L_n(z)$, а они совпадают со своими коммутативными образами. По этой причине неравенство влечёт совместность не только коммутативной, но и некоммутативной системы уравнений.

Наконец, отметим, что система уравнений (1) имеет бесконечно много решений, если множество её решений зависит хотя бы от одного произвольного ФСР от символов $x_1, \ldots, x_m$ [4].

Пример 1. Система из двух одинаковых уравнений $x_1z_1 - z_2x_2 = 0$ имеет бесконечно много решений, поскольку её решения можно записать в виде $z_1 = sx_2$, $z_2 = x_1s$, где s — произвольный ФСР от x_1, x_2.

3. Синтаксический анализ мономов методом мономиальных меток

Одной из важных проблем, связанных с разработкой систем и языков программирования, является проблема синтаксического анализа программ [2]. Как отмечено выше, большинство языков программирования является кс-языками, которые можно представить в виде ФСР, поэтому каждая программа, написанная на языке программирования, может рассматриваться как моном соответствующего ФСР. В связи с этим рассмотрим проблему синтаксического анализа мономов кс-языка.

Для того чтобы сформулировать её, рассмотрим подробнее систему полиномиальных уравнений Хомского — Щутценберже (2), которая определяет кс-язык. Как известно [1, 2], грамматика кс-языка является множеством правил подстановки

$$z_j \to q_{j1}(z,x),\ \ldots,\ z_j \to q_{jp_j}(z,x),\ j=1,\ldots,n, \tag{6}$$

где $q_{jk}(z,x)$ является мономом от некоммутативных символьных переменных с числовым коэффициентом, равным единице. Правила подстановки можно применять к начальному символу z_1, а затем к другим мономам в любом порядке неограниченное число раз, что позволяет выводить новые «правильные» мономы, образующие кс-язык. Известно также [1], что многочлены, стоящие в правой части системы уравнений Хомского — Щутценберже (2), определяются равенствами

$$Q_j(z,x) = \ q_{j1}(z,x) + \ldots + q_{jp_j}(z,x),\ j=1,\ldots,n.$$

Итак, проблема синтаксического анализа мономов состоит в том, чтобы определить, принадлежит ли моном данному кс-языку, т. е. может ли быть получен из начального символа z_1 при помощи правил подстановки (6), а также установить, какие правила подстановки и сколько раз использовались при выводе этого монома; при этом порядок использования правил подстановки не имеет значения [2].

В работе [8] предложен метод мономиальных меток, который позволяет провести беступиковый синтаксический анализ монома v от терминальных символов $x_1,\ldots,x_m$. Метод состоит в следующем. Сначала каждое правило подстановки $z_j \to q_{jk}(z,x)$ заменяется правилом $z_j \to t_{jk}q_{jk}(z,x)$, имеющим мономиальную метку t_{jk}, которая является символом из расширенного алфавита, и для новых правил вывода рассматривается соответствующая система уравнений Хомского — Щутценберже:

$$z_j = Q_j^*(z,x,t) \stackrel{\text{def}}{=} t_{j1}q_{j1}(z,x) + \ldots + t_{jp_j}q_{jp_j}(z,x),\ j=1,\ldots,n. \tag{7}$$

Далее решение системы (7) можно получить методом последовательных приближений, о котором говорилось в доказательстве теоремы 4:

$$z^{(k+1)}(x,t) = Q^*(z^{(k)}(x,t),x,t),\ k=0,1,\ldots;\ z^{(0)}=0.$$

Как результат, решение получается в виде ФСР

$$z_j = z_j^*(x,t) = \sum_{i=0}^{\infty}\langle z_j^*, w_i\rangle w_i,\ j=1,\ldots,n, \tag{8}$$

где w_i — мономы от символов $x_1,\ldots,x_m,t_{11},t_{12},\ldots,t_{np_n}$. Заметим, что если каждый символ t_{jk} заменить пустой цепочкой e, то решения систем уравнений (7) и (2) совпадают:

$$z_j^*(x,e) = z_j(x),\ j=1,\ldots,n. \tag{9}$$

Теперь итерации метода последовательных приближений для системы уравнений (7) дают многочлены возрастающей степени относительно всех символов $x_1, \ldots, x_m, t_{11}, t_{12}, \ldots, t_{np_n}$, при этом мономы степени не выше $\deg_x(v)$ относительно символов $x_1, \ldots, x_m$ после конечного числа итераций стабилизируются, не меняясь при последующих итерациях. Таким образом, можно получить начальные члены решения системы (7) в виде ФСР (8) до любой сколь угодно высокой степени, в том числе члены ФСР, представляющего первую компоненту этого решения:

$$z_1 = z_1^*(x,t) = \sum_{i=0}^{\infty} \langle z_j^*, w_i \rangle w_i. \tag{10}$$

Наконец, синтаксический анализ монома v кс-языка $z_1(x)$ можно провести следующим образом. Считывая мономы степени $\deg_x(v)$ относительно символов $x_1, \ldots, x_m$ и пропуская символы t_{jk}, можно установить, есть ли среди них моном v, а значит, можно ли вывести его с помощью системы продукций (6). При этом каждая мономиальная метка t_{jk}, содержащаяся в таком мономе, показывает, что при его выводе использовалось правило $z_j \to t_{jk} q_{jk}(z,x)$. В самом деле, из системы уравнений (7) и метода последовательных приближений нетрудно видеть, что, применяя это правило вывода к моному, мы умножаем его слева на символ t_{jk}. Следовательно, мономиальные метки монома решают проблему его синтаксического анализа, показывая, какие правила вывода кс-языка и сколько раз использовались при выводе этого монома, с точностью до порядка их применения.

Доказано [8], что метод мономиальных меток позволяет провести за конечное число шагов беступиковый синтаксический анализ любого монома (программы) кс-языка, заданного грамматикой (6).

Очевидно, что недостатком метода мономиальных меток является большое число громоздких итераций метода последовательных приближений, необходимых для получения начальных членов ФСР (8), причём это число возрастает вместе с ростом степени монома v. В связи с этим ниже предлагается другой путь для получения мономиальных меток некоммутативного монома.

4. Интегральное представление синтаксического многочлена

Информацию о мономиальных метках монома можно получить в виде $(n+m)$-кратного интеграла по циклу, где числа n и m не зависят от степени монома и равны числу нетерминальных и терминальных символов грамматики кс-языка соответственно.

Рассмотрим коммутативный образ ФСР (10)

$$\mathrm{ci}(z_1^*(x,t)) = \sum_{\alpha} s_\alpha(t) x^\alpha, \tag{11}$$

сгруппированный по степеням x^α в кратный ряд Гартогса.

Лемма 1. При всех мультииндеках α голоморфные в нуле коэффициенты ряда Гартогса $s_\alpha(t)$ являются многочленами.

Доказательство. Как отмечено выше, числовые коэффициенты системы уравнений (7) равны единице, поэтому коэффициенты ФСР (10) являются целыми положительными числами. Отсюда следует, что коэффициенты каждого степенного ряда также являются целыми положительными числами, а число $s_\alpha(e)$ равно сумме этих коэффициентов. Из равенств (9) и (11) следует равенство

$$\mathrm{ci}(z_1(x)) = \sum_{\alpha} s_\alpha(e) x^\alpha,$$

которое влечёт неравенство $s_\alpha(e) < \infty$. Это значит, что сумма неотрицательных целых коэффициентов степенного ряда функции $s_\alpha(e)$ конечна, следовательно, эта функция является многочленом. ■

Определение 1. *Синтаксическим многочленом монома v относительно кс-языка $z_1(x) = z_1^*(x, e)$* называется коэффициент $s_\alpha(t)$ ряда Гартогса (11), такой, что $x^\alpha = \mathrm{ci}(v)$.

Замечание 5. Мономиальные метки, содержащиеся в некоммутативных мономах кс-языка, не исчезают при переходе от ФСР (10) к его коммутативному образу (11) и сохраняются в виде мономов синтаксических многочленов, поскольку все коэффициенты ФСР (10) являются целыми положительными числами. Следовательно, если синтаксический многочлен монома относительно кс-языка равен нулю, то моном не принадлежит этому языку.

Замечание 6. Для проведения синтаксического анализа монома v, такого, что $\mathrm{ci}(v) = x^\alpha$, следует найти синтаксический многочлен $s_\alpha(t)$. Каждый моном многочлена $s_\alpha(t)$ является произведением мономиальных меток правил подстановки, которые позволяют получить некоторые мономы, имеющие тот же коммутативный образ x^α, поэтому для завершения синтаксического анализа монома v остаётся проверить, можно ли получить его с помощью правил подстановки, соответствующих всем мономам синтаксического многочлена $s_\alpha(t)$.

Следующая теорема даёт принципиальную возможность получить синтаксические многочлены $s_\alpha(t)$ в виде кратного интеграла по циклу, который может быть вычислен с помощью многомерных вычетов.

Теорема 5. При всех t, достаточно близких к нулю, и всех мультииндексах α синтаксический многочлен $s_\alpha(t)$ задаётся равенствами

$$s_\alpha(t) = \frac{1}{(2\pi i)^{n+m}} \int\limits_{\gamma_z \times \Gamma_x} \frac{z_1 \det\left(\delta_{ij} - (\mathrm{ci}(Q_i^*(z,x,t)))'_{z_j}\right)\, dz \wedge dx}{(z - \mathrm{ci}(Q^*(z,x,t)))\; x^{\alpha+I}}; \tag{12}$$

$$s_\alpha(t) = \frac{1}{(2\pi i)^n \alpha!} \int\limits_{\gamma_z} \frac{\partial^{\|\alpha\|}}{\partial x^\alpha} \left(\frac{z_1 \det\left(\delta_{ij} - (\mathrm{ci}(Q_i^*(z,x,t)))'_{z_j}\right)}{(z - \mathrm{ci}(Q^*(z,x,t)))} \right) \Big|_{x=0} dz, \tag{13}$$

где $\gamma_z = \{|z_1| = \ldots = |z_n|\}$ и $\Gamma_x = \{|x_1| = \cdots = |x_m|\}$ — циклы интегрирования; $0 < \delta \ll \varepsilon \ll 1$; $dz = dz_1 \wedge \ldots \wedge dz_n$; $dx = dx_1 \wedge \ldots \wedge dx_m$; $x^{\alpha+I} = x_1^{\alpha_1+1} \cdot \ldots \cdot x_m^{\alpha_m+1}$; $\alpha! = \alpha_1! \cdot \ldots \cdot \alpha_m!$; $(z - \mathrm{ci}(Q^*(z,x,t))) = (z_1 - \mathrm{ci}(Q_1^*(z,x,t))) \cdot \ldots \cdot (z_n - \mathrm{ci}(Q_n^*(z,x,t)))$; $\dfrac{\partial^{\|\alpha\|}}{\partial x^\alpha} = \dfrac{\partial^{\alpha_1+\ldots+\alpha_m}}{\partial x_1^{\alpha_1} \cdot \ldots \cdot \partial x_m^{\alpha_m}}$; δ_{ij} — символ Кронекера.

Для доказательства теоремы 5 понадобится следующая лемма.

Лемма 2. При всех (x,t), достаточно близких к нулю, голоморфная в нуле алгебраическая функция $\mathrm{ci}(z_1^*(x,t))$, представленная рядом Гартогса (11), задаётся формулой

$$\mathrm{ci}(z_1^*(x,t)) = \frac{1}{(2\pi i)^n} \int\limits_{\gamma_z} \frac{z_1 \det\left(\delta_{ij} - (\mathrm{ci}(Q_i^*(z,x,t)))'_{z_j}\right)}{(z - \mathrm{ci}(Q^*(z,x,t)))}\, dz. \tag{14}$$

Доказательство. Во-первых, отметим, что $\det(\delta_{ij} - (\mathrm{ci}(Q_i^*(z,x,t)))'_{z_j}) = J_1(z,x)$ — якобиан системы функций

$$\left(z_1 - \mathrm{ci}(Q_1^*(z,x,t)), \ldots, z_n - \mathrm{ci}(Q_n^*(z,x,t)\right)$$

по переменным $z_1, \ldots, z_n$. Так как $\det\left((\mathrm{ci}(Q_i^*(0,0,0)))'_{z_j}\right) = 0$, то $J_1(0,0) = 1$, следовательно, по теореме о неявном отображении система уравнений

$$z_j - \mathrm{ci}(Q_j^*(z,x,t)) = 0,\ j = 1, \ldots, n,$$

имеет в окрестности начала координат единственное и голоморфное решение:

$$z_j = \mathrm{ci}(z_j^*(x,t)) = \mathrm{ci}\left(\sum_{i=0}^{\infty} \langle z_j^*, w_j \rangle w_j\right),\ j = 1, \ldots, n.$$

Далее, если $z \in \gamma_z$, то при всех (x,t), достаточно близких к нулю, выполнены неравенства

$$|z_j| \geqslant |\mathrm{ci}(Q_j^*(z,x,t))|,\ j = 1, \ldots, n,$$

поэтому на цикле интегрирования знаменатель подынтегральной рациональной функции (14) не обращается в нуль. Наконец, из формулы многомерного логарифмического вычета [13] следует формула (14). ■

Доказательство теоремы 5. Коэффициенты ряда Гартогса, сходящегося в окрестности нуля, могут быть представлены формулой Коши

$$s_\alpha = \frac{1}{(2\pi i)^n} \int\limits_{\Gamma_x} \frac{\mathrm{ci}(z_1^*(x,t))\ dx}{x^{\alpha+I}} \tag{15}$$

либо формулой Тейлора

$$s_\alpha = \frac{1}{\alpha!} \frac{\partial^{\|\alpha\|}}{\partial x^\alpha} (\mathrm{ci}(z_1^*(0,t))).$$

Из этих формул и формулы (14) следует, что формулы (12) и (13) верны. ■

5. Пример вычисления синтаксического многочлена программы

Следуя [2], рассмотрим кс-язык арифметических выражений, порождённый грамматикой с системой правил подстановки

$$z_1 \to (z_1 * z_2),\ z_1 \to (a + b),\ z_2 \to b,$$

и выражение $(((a + b) * b) * b)$ как программу, написанную на этом языке. Известные методы синтаксического анализа показывают [2, с. 247], что для получения этой программы необходимо два раза использовать правило подстановки $z_1 \to (z_1 * z_2)$, один раз — правило $z_1 \to (a + b)$ и два раза — правило $z_2 \to b$.

Для того чтобы подчеркнуть, что интеграл (12), имеющий одну и ту же размерность $n + m$, позволяет анализировать мономы с неограниченными степенями, рассмотрим обобщённую программу, длина которой может быть сколь угодно большой.

Пример 2. Рассмотрим обобщённую программу $(\cdots((a + b) * b) \cdots * b)$, где символ «$*$» использован q раз и символ «$($» — $q + 1$ раз. Легко проверить, что программа содержит $4q + 5$ символов. По аналогии с предыдущей программой можно видеть, что обобщённую программу можно получить, используя первое и третье правила подстановки q раз, а второе правило один раз.

Сначала переобозначим терминальные символы «$+$», «$*$», «$($», «$)$», a, b символами x_1, x_2, x_3, x_4, x_5, x_6 соответственно, тогда система уравнений (7) для этого кс-языка примет вид

$$z_1 = t_{11} x_3 z_1 x_2 z_2 x_4 + t_{12} x_3 x_5 x_1 x_6 x_4,\ z_2 = t_{21} x_6. \tag{16}$$

Перепишем также эту программу как моном

$$x_3 \cdots x_3 x_3 x_5 x_1 x_6 x_4 x_2 x_6 x_4 \cdots x_2 x_6 x_4,$$

имеющий степень $4q+5$, которая может быть сколь угодно большой. Коммутативный образ этого монома равен

$$x_1 x_2^q x_3^{q+1} x_4^{q+1} x_5 x_6^{q+1}, \tag{17}$$

поэтому мультииндекс α в формуле (15) равен $(1, q, q+1, q+1, 1, q+1)$.

Далее, якобиан системы уравнений (16) в этом случае равен

$$\det(\delta_{ij} - (\mathrm{ci}(Q_i^*(z,x,t)))'_{z_j}) = 1 - t_{11}x_2x_3x_4z_2.$$

Теперь вычислим кратный интеграл (12) как повторный, сначала вычисляя интеграл (14), а затем (15). Интеграл (14) в нашем случае равен

$$\frac{1}{(2\pi i)^2} \int\limits_{|z_1|=|z_2|=\varepsilon} \frac{z_1(1 - t_{11}x_2x_3x_4z_2)\; dz_1 \wedge dz_2}{(z_1 - t_{11}x_2x_3x_4z_1z_2 - t_{12}x_1x_3x_4x_5x_6)(z_2 - t_{21}x_6)}. \tag{18}$$

Вычисляя последний интеграл по формуле Коши сначала по переменной z_2, а затем по z_1, получим

$$\begin{aligned}
\mathrm{ci}(z_1^*) &= \frac{1}{2\pi i} \int\limits_{|z_1|=\varepsilon} \frac{z_1(1 - t_{11}t_{21}x_2x_3x_4x_6)\; dz_1}{z_1 - t_{11}t_{21}x_2x_3x_4x_6z_1 - t_{12}x_1x_3x_4x_5x_6} = \\
&= \frac{1}{2\pi i} \int\limits_{|z_1|=\varepsilon} \frac{z_1(1 - t_{11}t_{21}x_2x_3x_4x_6)\; dz_1}{z_1(1 - t_{11}t_{21}x_2x_3x_4x_6) - t_{12}x_1x_3x_4x_5x_6} = \\
&= \frac{1}{2\pi i} \int\limits_{|z_1|=\varepsilon} \frac{z_1(1 - t_{11}t_{21}x_2x_3x_4x_6)\; dz_1}{\left(z_1 - \dfrac{t_{12}x_1x_3x_4x_5x_6}{1 - t_{11}t_{21}x_2x_3x_4x_6}\right)(1 - t_{11}t_{21}x_2x_3x_4x_6)} = \\
&= \frac{1}{2\pi i} \int\limits_{|z_1|=\varepsilon} \frac{z_1\; dz_1}{z_1 - \dfrac{t_{12}x_1x_3x_4x_5x_6}{1 - t_{11}t_{21}x_2x_3x_4x_6}} = \frac{t_{12}x_1x_3x_4x_5x_6}{1 - t_{11}t_{21}x_2x_3x_4x_6}.
\end{aligned}$$

Теперь вычислим интеграл (15), где $\alpha = (1, q, q+1, q+1, 1, q+1)$, $\Gamma_x = \{|x_1| = \ldots = |x_6|\}$, $t = (t_{11}, t_{12}, t_{21})$, используя разложение подынтегральной рациональной функции в степенной ряд для геометрической прогрессии:

$$\begin{aligned}
s_\alpha(t) &= \frac{1}{(2\pi i)^6} \int\limits_{\Gamma_x} \frac{t_{12}x_1x_3x_4x_5x_6\; dx_1 \wedge \cdots \wedge dx_6}{(1 - t_{11}t_{21}x_2x_3x_4x_6)\; x_1^2x_2^{q+1}x_3^{q+2}x_4^{q+2}x_5^2x_6^{q+2}} = \\
&= \frac{1}{(2\pi i)^6} \int\limits_{\Gamma_x} \left(\sum_{k=0}^{\infty} t_{11}^k t_{12} t_{21}^k x_1 x_2^k x_3^{k+1} x_4^{k+1} x_5 x_6^{k+1}\right) \frac{dx_1 \wedge \cdots \wedge dx_6}{x_1^2x_2^{q+1}x_3^{q+2}x_4^{q+2}x_5^2x_6^{q+2}} = \\
&= \sum_{k=0}^{\infty} \frac{t_{11}^k t_{12} t_{21}^k}{(2\pi i)^6} \int\limits_{\Gamma_x} x_2^{k-q}x_3^{k-q}x_4^{k-q}x_6^{k-q}\; \frac{dx_1 \wedge \cdots \wedge dx_6}{x_1 \ldots x_6} = t_{11}^q t_{12} t_{21}^q.
\end{aligned}$$

Очевидно, что интегралы от всех слагаемых равны нулю, кроме случая $k = q$, когда интеграл равен $(2\pi i)^6$. Итак, синтаксический многочлен мономов, имеющих коммутативный образ (17), равен $t_{11}^q t_{12} t_{21}^q$; согласно замечанию 6, мы должны проверить единственный путь для получения нашей программы: q раз использовать правило подстановки $z_1 \to (z_1 * z_2)$, один раз — правило $z_1 \to (a+b)$ и q раз — правило $z_2 \to b$, начиная

с исходного символа z_1. Этот путь в самом деле позволяет получить программу, что и завершает её синтаксический анализ.

Таким образом, пример 2 подтверждает, что интегральное представление фиксированной кратности даёт возможность осуществлять синтаксический анализ мономов (программ), степени (длины) которых неограничены.

ЛИТЕРАТУРА

1. *Salomaa A. and Soitolla M.* Automata-Theoretic Aspects of Formal Power Series. N.Y.: Springer Verlag, 1978. 167 p.
2. *Глушков В. М., Цейтлин Г. Е., Ющенко Е. Л.* Алгебра. Языки. Программирование. Киев: Наукова думка, 1973. 319 с.
3. *Гриффитс Ф., Харрис Дж.* Принципы алгебраической геометрии. Т. 1. М.: Мир, 1982. 259 с.
4. *Егорушкин О. И., Колбасина И. В., Сафонов К. В.* О совместности систем символьных полиномиальных уравнений и их приложении // Прикладная дискретная математика. Приложение. 2016. № 9. С. 119–121.
5. *Egorushkin O. I., Kolbasina I. V., and Safonov K. V.* On Solvability of Systems of Symbolic Polynomial Equations // Журнал Сибирского федерального университета. Математика и физика. 2016. Т. 9. № 2. С. 166–172.
6. *Семёнов А. Л.* Алгоритмические проблемы для степенных рядов и контекстно-свободных грамматик // Доклады Академии наук СССР. 1973. Т. 212. С. 50–52.
7. *Safonov K. V.* On power series of algebraic and rational functions in C^n // J. Math. Analysis and Appl. 2000. V. 243. P. 261–277.
8. *Сафонов К. В., Егорушкин О. И.* О синтаксическом анализе и проблеме В. М. Глушкова распознавания контекстно-свободных языков Хомского // Вестник Томского государственного университета. 2006. Приложение № 17. С. 63–67.
9. *Safonov K. V.* An algebraicity criterion for the sum of a power series (a generalization of Kronecker's criterion) and its application // Doklady Mathematics. 2009. V. 79. No. 1. P. 13–15.
10. *Pemantle R. and Wilson M. C.* Analytic Combinatorics in Several Variables. Cambridge: Cambridge University Press, 2013. 414 p.
11. *Herve M.* Several Complex Variables. Local Theory. Oxford: Oxford University, 1963. 158 p.
12. *Глухов М. М., Елизаров В. П., Нечаев А. А.* Алгебра. СПб.; М.; Краснодар: Лань, 2015. 608 с.
13. *Aizenberg L. A. and Yuzhakov A. P.* Integral Representations and Residues in Multidimensional Complex Analysis. Translations of Mathematical Monographs. V. 58. Providence: AMS, 1983. 283 p.

REFERENCES

1. *Salomaa A. and Soitolla M.* Automata-Theoretic Aspects of Formal Power Series. N.Y., Springer Verlag, 1978. 167 p.
2. *Glushkov V. M., Tseytlin G. E., and Yushchenko E. L.* Algebra. Yazyki. Programmirovanie [Algebra. Languages. Programming]. Kiev, Naukova dumka, 1973. 319 p. (in Russian)
3. *Griffiths P. and Harris J.* Principles of Algebraic Geometry. Vol. 1. N.Y., John Wiley and Sons, 1978. 348 p.
4. *Egorushkin O. I., Kolbasina I. V., and Safonov K. V.* O sovmestnosti sistem simvol'nykh polinomial'nykh uravneniy i ikh prilozhenii [On consistency of systems of symbolic polynomial equations and its application]. Prikladnaya Discretnaya Mathematika. Prilozhenie, 2016, no. 9, pp. 119–121. (in Russian)

5. *Egorushkin O. I., Kolbasina I. V., and Safonov K. V.* On solvability of systems of symbolic polynomial equations. J. Siberian Federal University. Mathematics and Physics, 2016, vol. 9, no. 2, pp. 166–172.
6. *Semenov A. L.* Algoritmicheskie problemy dlya stepennykh ryadov i kontekstno-svobodnykh grammatik [Algorithmic problems for power series and context-free grammars]. Doklady Akademii nauk SSSR, 1973, vol. 212, pp. 50–52. (in Russian)
7. *Safonov K. V.* On power series of algebraic and rational functions in C^n. J. Math. Analysis and Appl., 2000, vol. 243, pp. 261–277.
8. *Safonov K. V. and Egorushkin O. I.* O sintaksicheskom analize i probleme V. M. Glushkova raspoznavaniya kontekstno-svobodnykh yazykov Khomskogo [Syntactical analysis and the V. M. Glushkov problem of recognition of Chomsky Context-Free Languages]. Vestnik TSU. Prilozhenie, 2006, no. 17, pp. 63–67. (in Russian)
9. *Safonov K. V.* An algebraicity criterion for the sum of a power series (a generalization of Kronecker's criterion) and its application. Doklady Mathematics, 2009, vol. 79, no. 1, pp. 13–15.
10. *Pemantle R. and Wilson M. C.* Analytic Combinatorics in Several Variables. Cambridge, Cambridge University Press, 2013. 414 p.
11. *Herve M.* Several Complex Variables. Local Theory. Oxford, Oxford University, 1963. 158 p.
12. *Glukhov M. M., Elizarov V. P., and Nechaev A. A.* Algebra [Algebra]. SPb., Moscow, Krasnodar, LAN' Publ., 2015. 608 p. (in Russian)
13. *Aizenberg L. A. and Yuzhakov A. P.* Integral Representations and Residues in Multidimensional Complex Analysis. Translations of Mathematical Monographs. Vol. 58. Providence, AMS, 1983. 283 p.

УДК 32.973.26-018.2

ЭФФЕКТИВНАЯ ТРАНСЛЯЦИЯ ДЛЯ LL(1)-ГРАММАТИКИ НА ПРИМЕРЕ ЯЗЫКА ПРОГРАММИРОВАНИЯ

Ю. Л. Костюк

Национальный исследовательский Томский государственный университет, г. Томск, Россия

Предлагаются правила построения и функционирования транслятора для LL(1)-грамматики, генерирующего объектный язык. Транслятор представляется в виде таблицы ссылок на набор простых семантических программ. Таблица строится на основе порождающих правил грамматики, преобразованных в нестрогую нормальную форму Грейбах. Такой способ существенно упрощает разработку транслятора. Приведён пример построения транслятора и семантических программ для простого языка программирования, представленного обратной польской строкой.

Ключевые слова: *трансляция, LL(1)-грамматика, обратная польская строка, язык программирования.*

DOI 10.17223/20710410/37/7

EFFECTIVE TRANSLATION FOR LL(1)-GRAMMAR IN THE EXAMPLE OF A PROGRAMMING LANGUAGE

Yu. L. Kostyuk

Tomsk State University, Tomsk, Russia

E-mail: kostyuk_y_l@sibmail.com

Top-down analysis algorithms are the most useful for implementing translators for programming languages. The analysis methods based on the LL(1)-grammar assume that the generating rules of the grammar are previously transformed into a non-strict normal Greibach form. The transformed generating rules are written in an analyzer table. To convert an input string of characters to an object language, a special program for generating individual elements of the object language should be created for each generating rule. The development of the set of such programs is rather a complex and voluminous task when a translator is created. In this article, we propose a method for constructing semantic programs implementing the generation of elements of the object language when each semantic program is associated with the separate symbols in the generating rule but not with the generating rule as a whole. As a result, the semantic programs are extremely simple and convenient to be implemented. The method is described for an exemplifying programming language which is translated into a reverse Polish notation.

Keywords: *translation, LL(1) grammar, reverse Polish notation, programming language.*

Введение

К настоящему времени разработано большое количество различных алгоритмов грамматического анализа контекстно-свободных грамматик [1, 2]. Как отмечают многие исследователи, например Н. Вирт [3], наиболее удобными при реализации трансляторов для языков программирования являются алгоритмы анализа «сверху вниз», в частности метод рекурсивного спуска, когда каждое порождающее правило грамматики программируется в виде рекурсивной функции. Другой метод анализа «сверху вниз» на основе LL(1)-грамматики предполагает, что порождающие правила грамматики предварительно преобразуются в нестрогую нормальную форму Грейбах [2]. В такой форме все правые части порождающих правил либо пусты, либо начинаются с терминальных символов. Преобразованные порождающие правила заносятся в таблицу анализатора. Если, кроме анализа, требуется также преобразовать входную цепочку символов в объектный язык, то для каждого порождающего правила создается специальная программа генерации отдельных элементов объектного языка. Разработка набора таких программ является довольно сложной и объёмной задачей при создании транслятора. В данной работе предлагается такой способ построения семантических программ, реализующих генерацию элементов объектного языка, когда каждая программа связывается не с полным порождающим правилом, а с отдельными символами в порождающем правиле. В результате семантические программы получаются предельно простыми и удобными для реализации. В кратком виде эта идея была предложена автором в [4]. В данной работе такой способ изложен на примере простого языка программирования, который транслируется в обратную польскую строку (ОПС).

1. LL(1)-анализатор и транслятор для ОПС

Контекстно-свободная грамматика, преобразованная к нестрогой нормальной форме Грейбах, допускает детерминированный LL(1)-анализ, если для каждой группы порождающих правил с одним и тем же нетерминальным символом в левой части разные правые части будут однозначно различимы по первому терминальному символу. Пусть входная цепочка символов всегда заканчивается символом-ограничителем «$\perp$». Для работы LL(1)-анализатора необходимо построить таблицу, в которой столбцы помечены терминальными символами (включая ограничитель), а строки — нетерминальными символами преобразованной грамматики. Для всех порождающих правил грамматики, имеющих вид

$$A \to a\gamma$$

(где A — нетерминальный символ; a — терминальный символ; γ — цепочка из терминальных и нетерминальных символов), правая часть правила $a\gamma$ заносится на пересечение строки, помеченной символом A, и столбца, помеченного символом a. Если порождающее правило имеет вид

$$A \to \lambda,$$

где λ — пустая цепочка, то во все клетки строки, помеченной нетерминальным символом A и не занятые другими правилами, записывается λ.

В качестве примера рассмотрим грамматику простых арифметических выражений [1], в которой символы S, T, F — нетерминальные, причём S — начальный. Символы «+», «*», круглые скобки и a — терминальные; вертикальная черта разделяет различные правые части при одном и том же нетерминальном символе в левой части правил:

$$S \to S + T \mid T$$
$$T \to T * F \mid F$$
$$F \to (S) \mid a$$

После преобразования грамматики к нестрогой нормальной форме Грейбах получим порождающие правила с двумя дополнительными нетерминальными символами U и V:

$$S \to (S)VU \mid aVU$$
$$U \to +TU \mid \lambda$$
$$T \to (S)V \mid aV$$
$$V \to *FV \mid \lambda$$
$$F \to (S) \mid a$$

Функционирование анализатора. До начала работы в стек анализатора записывается ограничитель «$\perp$» (считающийся терминальным символом), а затем начальный нетерминальный символ. На каждом шаге анализатор проверяет, допустим ли очередной входной символ, и если да, то выполняет одно из двух действий:

1) если на вершине стека нетерминальный символ, то, в зависимости от того, каков очередной входной символ, этот нетерминальный символ заменяется в стеке символами правой части соответствующего правила, причём символы записываются в обратном порядке. Если для очередного входного символа в таблице записано λ, то нетерминальный символ просто удаляется из стека, а если в таблице пустая клетка, то анализатор фиксирует ошибку во входной цепочке;
2) если на вершине стека терминальный символ, то он сравнивается с очередным входным символом. При совпадении терминальный символ удаляется из стека и делается переход к следующему символу входной цепочки. При несовпадении анализатор фиксирует ошибку.

Работа анализатора прекращается, когда входная цепочка символов окажется полностью просмотренной. Если при этом стек пуст, то цепочка считается правильной, если не пуст — ошибочной.

Трансляция входной цепочки символов в ОПС выполняется параллельно с работой LL(1)-анализатора. Для транслятора необходим второй стек, действия с которым выполняются синхронно с действиями со стеком анализатора. Во второй стек записываются семантические символы, которые генерируют элементы ОПС. Генерация элементов ОПС выполняется при извлечении семантических символов из второго стека.

Для реализации транслятора наряду с основной таблицей LL(1)-анализатора необходимо задать семантическую таблицу. Её размеры в точности совпадают с таблицей анализатора, более того, для каждой непустой клетки таблицы анализатора, где находится правая часть какого-либо правила порождения, в семантическую таблицу помещаются семантические символы, количество которых в точности такое же, как в правой части этого правила порождения. При этом надо учесть, что терминальные символы в контекстно-свободной грамматике на самом деле являются лексемами, распознаваемыми лексическим анализатором. Некоторые из лексем (имена переменных и константы) содержат дополнительную семантическую информацию, т. е. одинаковые терминальные символы в контекстно-свободной грамматике могут различаться семантически.

В качестве примера рассмотрим семантические действия для грамматики простых арифметических выражений, обозначаемые следующими семантическими символами:

a — запись в ОПС операнда из входной цепочки (переменной или константы);

«+» — запись в ОПС операции сложения;

«∗» — запись в ОПС операции умножения;

«○» — пустое действие.

Действия выполняются в момент выталкивания соответствующих символов из второго стека. В табл. 1 представлена совмещённая таблица: анализатора и семантическая таблица транслятора.

Таблица 1

Нетерм. символ	$+$	$*$	$($	$)$	a	$\perp$
S			$(S)VU$ ○ ○ ○ ○ ○		aVU a ○ ○	
U	$+TU$ ○ ○ $+$	λ	λ	λ	λ	λ
T			$(S)V$ ○ ○ ○ ○		aV a ○	
V	λ	$*FV$ ○ ○ $*$	λ	λ	λ	λ
F			(S) ○ ○ ○		a a	

Так как одинаковые терминальные символы a в контекстно-свободной грамматике семантически могут различаться, во входной цепочке различные операнды будем обозначать разными символами: x, y и др., при этом все они соответствуют одному и тому же терминальному символу a. При этом операнды могут быть различных типов — именами переменных, константами, — а лексический анализатор, непосредственно просматривающий входную цепочку символов, на каждом шаге работы LL(1)-анализатора выдаёт ему тип терминального символа (лексемы).

До начала работы в стек анализатора записывается символ-ограничитель и начальный нетерминальный символ, а во второй стек — два символа, обозначающих пустое действие. На каждом шаге работы транслятора делается следующее:

— если верхний символ в стеке анализатора нетерминальный, то во втором стеке верхний символ (который при этом будет обязательно пустым действием) удаляется и во второй стек записывается последовательность символов из семантической таблицы (эта последовательность находится на пересечении строки, помеченной нетерминалом, и столбца, помеченного входным терминалом);

— если верхний символ в стеке анализатора — терминал, то его тип должен соответствовать типу терминала на входе (в противном случае — ошибка), и тогда из второго стека удаляется верхний символ и выполняется соответствующее действие; если это действие — операция, то она просто записывается в ОПС, а если это символ a и терминал на входе — константа, то значение этой константы записывается в ОПС; если действие — символ a и терминал на входе — имя, то генератор ищет это имя в таблице переменных, и если оно там найдено, то в ОПС записывается тип переменной и ссылка на неё в таблице.

В последнем случае, если имя в таблице переменных отсутствует, то действия зависят от входного языка. Если в языке требуется, чтобы все имена были предварительно

описаны, то фиксируется ошибка, а если имена считаются описанными по умолчанию, то в таблицу переменных заносится новое имя.

Пример работы LL(1)-анализатора и генератора ОПС, заданного в табл. 1, для входной цепочки «$x * (c + d)\ \perp$» представлен в табл. 2.

Таблица 2

№ шага	Входные символы	Содержимое стека анализатора	Содержимое второго стека	Порождающее правило	ОПС
1	$x * (c + d)\ \perp$	$S\ \perp$	$\circ\ \circ$	$S \to aVU$	
2	$x * (c + d)\ \perp$	$aVU\ \perp$	$a\ \circ\ \circ\ \circ$		x
3	$*(c + d)\ \perp$	$VU\ \perp$	$\circ\ \circ\ \circ$	$V \to *FV$	x
4	$*(c + d)\ \perp$	$*FVU\ \perp$	$\circ\ \circ * \circ\ \circ$		x
5	$(c + d)\ \perp$	$FVU\ \perp$	$\circ * \circ\ \circ$	$F \to (S)$	x
6	$(c + d)\ \perp$	$(S)VU\ \perp$	$\circ\ \circ\ \circ * \circ\ \circ$		x
7	$c + d)\ \perp$	$S)VU\ \perp$	$\circ\ \circ * \circ\ \circ$	$S \to aVU$	x
8	$c + d)\ \perp$	$aVU)VU\ \perp$	$a\ \circ\ \circ\ \circ * \circ\ \circ$		$x\ c$
9	$+d)\ \perp$	$VU)VU\ \perp$	$\circ\ \circ\ \circ * \circ\ \circ$	$V \to \lambda$	$x\ c$
10	$+d)\ \perp$	$U)VU\ \perp$	$\circ\ \circ + \circ * \circ\ \circ$	$U \to +TU$	$x\ c$
11	$+d)\ \perp$	$+TU)VU\ \perp$	$\circ\ \circ + \circ * \circ\ \circ$		$x\ c$
12	$d)\ \perp$	$TU)VU\ \perp$	$\circ + \circ * \circ\ \circ$	$T \to aV$	$x\ c$
13	$d)\ \perp$	$aVU)VU\ \perp$	$a\ \circ + \circ * \circ\ \circ$		$x\ c\ d$
14	$)\ \perp$	$VU)VU\ \perp$	$\circ + \circ * \circ\ \circ$	$V \to \lambda$	$x\ c\ d$
15	$)\ \perp$	$U)VU\ \perp$	$+\ \circ * \circ\ \circ$	$U \to \lambda$	$x\ c\ d+$
16	$)\ \perp$	$)VU\ \perp$	$\circ * \circ\ \circ$		$x\ c\ d+$
17	$\perp$	$VU\ \perp$	$*\ \circ\ \circ$	$V \to \lambda$	$x\ c\ d+*$
18	$\perp$	$U\ \perp$	$\circ\ \circ$	$U \to \lambda$	$x\ c\ d+*$
19	$\perp$	$\perp$	$\circ$		$x\ c\ d+*$

2. Простой язык програмирования и ОПС для него

Рассмотрим действия транслятора на примере простого языка программирования, в котором определены почти все типичные для таких языков элементы. Определим синтаксис языка следующим набором порождающих правил:

$P \to$ **begin** DAB **end**
$D \to$ **dim** $a[k]; D \mid \lambda$
$A \to aH = S \mid$ **if** C **then** $ABE \mid$ **while** C **do** AB **end** $\mid$ **in** $aH \mid$ **out** S
$B \to; AB \mid \lambda$
$E \to$ **else** AB **end** $\mid$ **end**
$C \to S = S \mid S < S \mid S > S \mid S \neq S$
$S \to S + T \mid S - T \mid T$
$T \to T * F \mid T/F \mid F$
$F \to (S) \mid aH \mid k$
$H \to [S] \mid \lambda$

Нетерминальные символы, обозначенные заглавными латинскими буквами, имеют следующий смысл:

— P — начальный нетерминальный символ, порождает всю программу;
— D — определяет описания массивов;
— A — определяет операторы (присваивание, условный, цикл, ввод, вывод);

— B — определяет последовательность операторов;
— E — определяет вторую часть условного оператора;
— C — определяет сравнение (равно, меньше, больше, не равно) для двух выражений;
— S, T, F — определяют выражение с операциями сложения, вычитания, умножения, деления, а также скобками;
— H — определяет индексирование массива.

Жирным шрифтом выделены зарезервированные служебные слова, они, наряду со знаками операций, скобками, именами (обозначаемыми буквой a) и числовыми константами (обозначаемыми буквой k), являются терминальными символами для контекстно-свободной грамматики и распознаются лексическим анализатором как лексемы. Имена простых переменных в языке описаны по умолчанию; в описании массива константа k задаёт количество элементов в нём. Типы переменных, массивов и констант в общем случае вещественные.

При трансляции генерируется не только ОПС, но также таблица массивов и таблица простых переменных. Каждый элемент таблицы массивов содержит имя массива, длину и ссылку на начало расположения его элементов в памяти. Каждый элемент таблицы простых переменных содержит имя и ссылку на расположение этой переменной в памяти. Эти таблицы используются в процессе работы транслятора.

Отдельный элемент ОПС может быть операндом или операцией. Интерпретация (исполнение) ОПС как программы производится с использованием стека для операндов и результатов операций. Операнды последовательно заносятся в стек. Если встретилась операция, то она исполняется с использованием операндов в верхней части стека, их количество определяется операцией. Результат операции либо заносится в стек, либо запоминается другим способом.

Операнд в ОПС состоит из двух частей: типа, а также значения или ссылки. Типы операндов:

— переменная, ссылка указывает расположение переменной в памяти;
— массив, ссылка указывает расположение начального элемента массива в памяти;
— константа, вторая часть — её значение;
— метка, ссылка указывает порядковый номер того элемента ОПС, на который требуется передать управление.

Операция в ОПС задаётся её номером или обозначением. Для рассматриваемого языка определены следующие операции:

— арифметические (сложение, вычитание, умножение, деление), обозначаемые знаками «+», «−», «*», «/»; они требуют по два операнда, результат заносится в стек;
— сравнения (равно, меньше, больше, не равно), обозначаемые знаками «=», «<», «>», «≠»; они требуют по два операнда, результат может быть 1 (истина) или 0 (ложь) и заносится в стек;
— присваивание, обозначается знаком «←» (чтобы различать с операцией равно); требует два операнда, результат (значение второго операнда) запоминается в памяти по ссылке (из первого операнда), в стек ничего не заносится;
— индексация, обозначается **ind**, требует два операнда, ссылка на массив (первый операнд) складывается со значением второго операнда (индексом), результат (ссылка на индексируемый элемент массива) заносится в стек; предполагается, что элементы в массиве пронумерованы числами 0, 1 и т. д.;

— ввод, обозначается **in**, требует один операнд — ссылку, вводит числовое значение со стандартного устройства ввода и запоминает в памяти по ссылке, в стек ничего не заносится;
— вывод, обозначается **out**, требует один операнд — значение, выводит это значение на стандартное устройство вывода в виде изображения числа, в стек ничего не заносится;
— выделение памяти для массивов и простых переменных, обозначается **mem**, требует один операнд, выделяет память указанного размера, в стек ничего не заносится;
— освобождение выделенной памяти, обозначается **del**, не требует операндов, в стек ничего не заносится;
— безусловный переход по метке, обозначается **j**, требует один операнд — метку (номер элемента ОПС), производит переход к указанному элементу ОПС, в стек ничего не заносится;
— условный переход по метке, обозначается **jf**, требует два операнда; если значение первого операнда равно 0 (ложь), то производит переход по второму операнду — метке — к указанному элементу ОПС; в противном случае ничего не делается; в обоих случаях в стек ничего не заносится.

3. Трансляция языка програмирования

Для построения транслятора необходимо сначала преобразовать порождающие правила языка в нестрогую нормальную форму Грейбах. Если после этого для некоторого нетерминального символа различные правые части начинаются с одного и того же терминального символа, то к ним можно применить преобразование, называемое факторизацией. При этом в грамматику добавится ещё один нетерминальный символ.

Затем для каждой правой части порождающего правила (исключая пустое порождение λ) необходимо сконструировать последовательность семантических действий, генерирующих элементы ОПС, и записать их в виде семантических символов. Их должно быть ровно столько, сколько символов (терминальных и нетерминальных) имеется в соответствующей правой части порождающего правила. Семантические символы обозначают следующие действия:

— a — имя, распознанное лексическим анализатором из входной цепочки символов, ищется в таблице переменных и в таблице массивов; если отсутствует в обеих таблицах, то добавляется в таблицу переменных, при этом счетчик количества занятой памяти увеличивается на 1; затем ссылка на расположение в памяти этой переменной или массива записывается в ОПС как операнд;
— k — значение константы, распознанное лексическим анализатором из входной цепочки символов, записывается в ОПС как операнд;
— операция, обозначенная знаком («+», «−» и др.) или словом (**in**, **out** и др.) — записывается в ОПС;
— «○» — пустое действие;
— число (от 1 до 9) — задаёт выполнение соответствующей семантической программы, алгоритмы этих программ приведены ниже.

Для синхронизации работы транслятора с первым и вторым стеком некоторые порождающие правила в конце правой части необходимо дополнить нетерминальным символом Z, который всегда порождает пустую цепочку. Например, правило $A \to a = S$ расширено до $A \to a = SZ$. Это сделано для того, чтобы операция присваивания записывалась в ОПС после того, как будут сгенерированы все операнды и операции в правой части присваивания.

Далее приведены преобразованные порождающие правила и взятые в фигурные скобки соответствующие им последовательности семантических символов:

$P \to$ **begin** DAB **end** $\{1 \circ \circ \circ 9\}$
$D \to$ **dim** $a[k]; D$ $\{\circ\, 2 \circ 3 \circ \circ \circ\} \mid \lambda$
$A \to aH = S\ \{a \circ \circ \circ \leftarrow\} \mid$ **if** C **then** $ABE\ \{\circ \circ 4 \circ \circ \circ\}$ $\mid$
while C **do** AB **end** $\{5 \circ 4 \circ \circ 6\} \mid$ **in** $aHZ\{\circ \circ \circ$ **in**$\} \mid$ **out** $SZ\{\circ \circ$ **out**$\}$
$B \to ; AB\ \{\circ \circ \circ\} \mid \lambda$
$E \to$ **else** AB **end** $\{7 \circ \circ 8\} \mid$ **end** $\{8\}$
$C \to (S)VUQ\ \{\circ \circ \circ \circ \circ \circ \circ\} \mid aVUQ\ \{\circ \circ \circ \circ\} \mid kVUQ\ \{\circ \circ \circ \circ\}$
$Q \to = SZ\ \{\circ \circ =\} \mid < SZ\ \{\circ \circ <\} \mid > SZ\ \{\circ \circ =\} \mid \neq SZ\ \{\circ \circ \neq\}$
$S \to (S)VU\ \{\circ \circ \circ \circ \circ \circ\} \mid aVU\ \{\circ \circ \circ\} \mid kVU\ \{\circ \circ \circ\}$
$U \to +TU\ \{\circ \circ +\} \mid -TU\ \{\circ \circ -\} \mid \lambda$
$T \to (S)V\ \{\circ \circ \circ \circ \circ\} \mid aV\ \{\circ \circ\} \mid kV\ \{\circ \circ\}$
$V \to *FV\ \{\circ \circ *\} \mid /FV\ \{\circ \circ /\} \mid \lambda$
$F \to (S)\ \{\circ \circ \circ\} \mid aH\ \{a \circ\} \mid k\ \{k\}$
$H \to [S]\ \{\circ \circ$ **ind**$\} \mid \lambda$
$Z \to \lambda$

Рассмотрим алгоритмы семантических программ. При их выполнении используется ещё один, третий, стек, а также две переменные — i и m. Переменная i — счётчик (количество) сгенерированных элементов ОПС, m — размер выделенной памяти для массивов и переменных. Счётчик i увеличивается на 1 всякий раз, как только генерируется очередной элемент ОПС.

Программа 1.

1. $i := 1$, $m := 0$.
2. В ОПС записывается 0 — место для записи в последующем размера выделенной памяти.
3. В третий стек записывается 1 (ссылка на зарезервированный элемент ОПС).

Программа 2.

1. Имя, распознанное лексическим анализатором из входной цепочки символов, ищется в таблице массивов.
2. Если имя найдено, то фиксируется ошибка.
3. Если имя в таблице массивов отсутствует, то добавляется в таблицу массивов.
4. В ту же строку таблицы массивов записывается значение переменной m (ссылка на начало расположения массива в памяти).

Программа 3.

1. Константа, распознанная лексическим анализатором из входной цепочки символов, добавляется к переменной m.

Программа 4.

1. В третий стек записывается i.
2. В ОПС записывается 0 — место для будущей метки.
3. В ОПС записывается операция **jf** — переход при условии `false`.

Программа 5.

1. В третий стек записывается i.

Программа 6.

1. Через верхний элемент третьего стека, как ссылку на ранее заготовленное место для метки, записывается $i + 2$.

2. В ОПС записывается метка, значение для которой читается из третьего стека.
3. В ОПС записывается операция **j** — безусловный переход.

Программа 7.

1. Через верхний элемент третьего стека, как ссылку на ранее заготовленное место для метки, записывается $i + 2$.
2. В третий стек записывается i.
3. В ОПС записывается 0 — место для будущей метки.
4. В ОПС записывается операция **j** — безусловный переход.

Программа 8.

1. Через верхний элемент третьего стека, как ссылку на ранее заготовленное место для метки, записывается i.

Программа 9.

1. Через верхний элемент третьего стека, как ссылку на ранее заготовленное место в ОПС, записывается значение переменной m (размер выделенной памяти для массивов и переменных).

Благодаря использованию третьего стека генерируется корректная ОПС для вложенных условных операторов и циклов. Например, по следующей входной цепочке символов (фрагменту программы, которая упорядочивает массив `A`):

```
i=1;
while i<n do
  if A[i-1]>A[i] then
    x=A[i-1]; A[i-1]=A[i]; A[i]=x;
    if i>1 then i=i-1 end
  else i=i+1 end
end;
```

будет сгенерирована следующая ОПС, в которой метки как операнды обозначены именами `m1, m2, m3,` а места ОПС, куда они ссылаются, вынесены влево и обозначены именами меток с двоеточием:

```
    i 1 ←
m1: i n < m3 jf
    A i 1 - ind A i ind > m2 jf
    x A i 1 - ind ← A i 1 - ind A i ind ← A i ind x ←
    i 1 > m2 jf i i 1 - ←
m2: i i 1 + ←
    m1 j
m3:
```

Заключение

Рассмотренный пример грамматики простейшего языка программирования является демонстрационным, но его нетрудно расширить, добавив и другие операторы, операции, типы и структуры переменных и констант, описания и т. п., получив в результате универсальный язык программирования. Использование предложенных принципов построения транслятора позволяет большую часть описания анализируемого языка разместить в таблицах, за исключением некоторых действий, программируемых дополнительно в виде небольших семантических программ. В результате трудозатраты при разработке транслятора существенно сокращаются. Использование ОПС не является существенным ограничением, так как это удобная форма внутреннего представ-

ления транслируемой программы. При необходимости, скорректировав правила функционирования транслятора, можно генерировать иные структуры (тройки, четвёрки и т. п.). Наконец, на основе полученной ОПС можно генерировать последовательность машинных команд.

ЛИТЕРАТУРА

1. *Ахо А. В., Лам М. С., Сети Р., Ульман Дж. Д.* Компиляторы: принципы, технологии и инструментарий. 2-е изд. М.: Вильямс, 2008. 1184 с.
2. *Mogensen T. A.* Introduction to Compiler Design. Springer, 2011. 225 p.
3. *Вирт Н.* Построение компиляторов. М.: ДМК Пресс, 2010. 192 с.
4. *Костюк Ю.Л̇.* Табличный способ генерации обратной польской строки для LL(1)-грамматики // Информационные технологии и математическое моделирование (ИТММ-2015). Материалы XIV Междунар. конф. им. А. Ф. Терпугова, 18–22 ноября 2015. Ч. 1. Томск: Изд-во Том. ун-та, 2015. С. 190–195.

REFERENCES

1. *Aho V. A., Lam M. S., Sethi R., and Ullman J. D.* Compilers: Principles, Techniques, and Tools. Boston, Pearson Education, 2007.
2. *Mogensen T. A.* Introduction to Compiler Design. Springer, 2011. 225 p.
3. *Virt N.* Theory and Techniques of Compiler Construction. Addison-Wesley, Reading, 1996.
4. *Kostyuk Yu. L.* Tablichnyj sposob generacii obratnoj pol'skoj stroki dlya LL(1)-grammatiki [The tabular method of generating reverse Polish notation for LL(1) grammars]. Proc. ITMM-2015, vol. 1. Tomsk, TSU Publ., 2015, pp. 190–195. (in Russian)

УДК 510.52

О ГЕНЕРИЧЕСКОЙ СЛОЖНОСТИ ПРОБЛЕМЫ РАЗРЕШИМОСТИ СИСТЕМ ДИОФАНТОВЫХ УРАВНЕНИЙ В ФОРМЕ СКОЛЕМА[1]

А. Н. Рыбалов

Омский государственный университет им. Ф. М. Достоевского, г. Омск, Россия,
Институт математики им. С. Л. Соболева СО РАН, г. Новосибирск, Россия

Изучается генерическая сложность десятой проблемы Гильберта для систем диофантовых уравнений в форме Сколема. Приводится генерический полиномиальный алгоритм, определяющий разрешимость таких систем уравнений над множеством натуральных чисел (без нуля). Доказывается, что проблема разрешимости таких систем уравнений над множеством целых чисел является неразрешимой на любом рекурсивном строго генерическом подмножестве входов. Доказательство этой теоремы проходит также для случая, когда решения ищутся во множестве натуральных чисел с нулём.

Ключевые слова: *генерическая сложность, диофантовы уравнения.*

DOI 10.17223/20710410/37/8

ON GENERIC COMPLEXITY OF DECIDABILITY PROBLEM FOR DIOPHANTINE SYSTEMS IN THE SKOLEM'S FORM

A. N. Rybalov

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

E-mail: alexander.rybalov@gmail.com

Generic-case approach to algorithmic problems was suggested by Miasnikov, Kapovich, Schupp and Shpilrain in 2003. This approach studies behavior of an algorithm on typical (almost all) inputs and ignores the rest of inputs. This approach has applications in cryptography where it is required that algorithmic problems must be difficult for almost all inputs. Romankov in 2012 shows that the basic encryption functions of many public key cryptographic systems, among which the RSA system and systems, based on the intractability of the discrete logarithm problem, can be written in the language of Diophantine equations. The effective generic decidability of these equations leads to hacking of corresponding systems, therefore it is actual to study the generic complexity of the decidability problem for Diophantine equations in various formulations. For example, Rybalov in 2011 proved that the Hilbert's tenth problem remains undecidable on strongly generic subsets of inputs in the representation of Diophantine equations by so-called arithmetic schemes. In this paper, we study generic complexity of the Hilbert's tenth problem for systems of Diophantine equations in the Skolem's form. We construct generic polynomial algorithm for determination of solvability of such systems over natural numbers (without zero). We prove strongly generic undecidability of this problem for systems over integers and over natural numbers with zero.

[1]Работа поддержана грантом РФФИ № 15-41-04312.

Keywords: *generic complexity, diophantine equations.*

Введение

В 1900 г. Д. Гильберт на II математическом конгрессе сформулировал 23 математические проблемы, которые XIX век оставил в наследство XX веку. Десятая проблема в современной формулировке звучит следующим образом: найти алгоритм, который по любому диофантовому уравнению определяет, разрешимо ли оно в целых числах. В 1970 г. Ю. В. Матиясевич [1], основываясь на работах М. Дэвиса, Дж. Робинсон и Х. Патнема, доказал, что такого алгоритма не существует. В дальнейшем было доказано, что десятая проблема Гильберта алгоритмически неразрешима уже для диофантовых уравнений от тринадцати неизвестных (Матиясевич, Робинсон [2]). Оценка улучшена до девяти неизвестных (Дж. Джонс [3]).

Генерический подход в применении к алгоритмическим проблемам впервые предложен в 2003 г. в [4]. В рамках этого подхода изучается поведение алгоритма на множестве «почти всех» входов (это множество называется генерическим), игнорируя поведение на остальных входах, на которых алгоритм может работать медленно или вообще не останавливаться. Такой подход имеет приложение в криптографии, где требуется, чтобы алгоритмические проблемы были трудными для «почти всех» входов. В [5 – 7] показано, что основные функции шифрования многих криптографических систем с открытым ключом, среди которых система RSA и системы, основанные на трудноразрешимости проблемы дискретного логарифма, записываются на языке диофантовых уравнений. Эффективная генерическая разрешимость этих уравнений приводит к взлому соответствующих систем, поэтому актуальной является задача изучения генерической сложности проблемы разрешимости диофантовых уравнений в различных её постановках. Например, в [8, 9] доказано, что десятая проблема Гильберта остаётся неразрешимой на строго генерических подмножествах входов при представлении диофантовых уравнений с помощью так называемых арифметических схем. В данной работе изучается генерическая сложность десятой проблемы Гильберта для систем диофантовых уравнений в форме Сколема. Приводится генерический полиномиальный алгоритм для проблемы разрешимости таких систем уравнений над множеством натуральных чисел (без нуля). Доказывается, что проблема разрешимости таких систем уравнений над множеством целых чисел является неразрешимой на любом рекурсивном строго генерическом подмножестве входов.

1. Генерические алгоритмы

Пусть I — некоторое множество. Через $\mathbb{N}$ будем обозначать множество натуральных чисел (без нуля), а через $\mathbb{N}_0$ — натуральные числа с нулём. Функция $\mathrm{size} : I \to \mathbb{N}_0$ называется *функцией размера*, если для любого $n \in \mathbb{N}_0$ множество $I_n = \{x \in I : \mathrm{size}(x) = n\}$ конечно. Например, на множестве $I = \Sigma^*$ слов конечного алфавита Σ естественно определяется функция размера (длины) $w \mapsto |w|$, где $|w|$ — длина слова w. На $\mathbb{N}$ также естественно определяется функция размера (длины двоичной записи) $n \mapsto |n|_2$, где $|n|_2$ — длина двоичной записи числа n. Как обычно делается в теории вычислимости, будем под алгоритмическими проблемами понимать проблемы распознавания подмножеств из некоторого множества входов с определённой на нём функцией размера. Зафиксируем некоторое множество входов I и функцию размера на этом множестве. Для подмножества $S \subseteq I$ определим последовательность

$$\rho_n(S) = \frac{|S_n|}{|I_n|},\ n = 1, 2, 3, \ldots,$$

где $S_n = S \cap I_n$ — множество входов из S размера n. Здесь $|A|$ — число элементов в множестве A. Заметим, что $\rho_n(S)$ — это вероятность попасть в S при случайной и равновероятной генерации входов из I_n. *Асимптотической плотностью* S назовём предел (если он существует)

$$\rho(S) = \lim_{n\to\infty} \rho_n(S).$$

Множество S называется *генерическим*, если $\rho(S) = 1$, и *пренебрежимым*, если $\rho(S) = 0$. Очевидно, что S генерическое тогда и только тогда, когда его дополнение $I \setminus S$ пренебрежимо.

Следуя [4], назовём множество S *строго пренебрежимым*, если последовательность $\rho_n(S)$ экспоненциально быстро сходится к 0, т. е. существуют константы $0 < \sigma < 1$ и $C > 0$, такие, что для любого n

$$\rho_n(S) < C\sigma^n.$$

Теперь S называется *строго генерическим*, если его дополнение $I \setminus S$ строго пренебрежимо.

Алгоритм $\mathcal{A}$ с множеством входов I и множеством выходов $J \cup \{?\}$ $(? \notin J)$ называется (*строго*) *генерическим*, если

1) $\mathcal{A}$ останавливается на всех входах из I;
2) множество $\{x \in I : \mathcal{A}(x) = ?\}$ является (строго) пренебрежимым.

Здесь через $\mathcal{A}(x)$ обозначается результат работы алгоритма $\mathcal{A}$ на входе x. Генерический алгоритм $\mathcal{A}$ *вычисляет функцию* $f : I \to J$; если для $x \in I$ не выполнено $\mathcal{A}(x) = ?$, то $f(x) = \mathcal{A}(x)$. Ситуация $\mathcal{A}(x) = ?$ означает, что $\mathcal{A}$ не может вычислить функцию f на аргументе x. Но условие 2 гарантирует, что $\mathcal{A}$ корректно вычисляет f на почти всех входах (входах из генерического множества). Множество $S \subseteq I$ называется (*строго*) *генерически разрешимым*, если существует (строго) генерический алгоритм, вычисляющий его характеристическую функцию.

2. Системы диофантовых уравнений в форме Сколема

Система диофантовых уравнений записана в *форме Сколема* [10], если каждое уравнение в ней имеет один из следующих типов:

1) $x_i = x_j x_k$,
2) $x_i = x_j + x_k$,
3) $x_i = 1$.

Нетрудно показать, что для любого диофантова уравнения (системы) можно эффективно построить эквивалентную ему систему диофантовых уравнений в форме Сколема. Например, для уравнения $x^2 - 3y^2 = 1$ эквивалентной системой Сколема является

$$x_1 = x_2x_2,\ x_3 = x_4x_4,\ x_5 = x_3 + x_3,\ x_6 = x_5 + x_3,\ x_7 = 1,\ x_8 = x_6 + x_7,\ x_8 = x_1.$$

Будем называть систему в форме Сколема *нормализованной*, если в k-м уравнении системы могут встречаться только переменные x_i, где $i \leqslant 3k$. Очевидно, что любую систему в форме Сколема можно нормализовать при помощи подходящего перенумерования переменных. В дальнейшем будем рассматривать только нормализованные системы диофантовых уравнений в форме Сколема. Размер такой системы — это число уравнений в ней. Обозначим через $\mathcal{S}$ множество всех нормализованных систем в форме Сколема.

Лемма 1. Число нормализованных систем в форме Сколема размера n есть

$$|\mathcal{S}_n| = \prod_{k=1}^{n}(54k^3 + 3k).$$

Доказательство. Для k-го уравнения в системе $S \in \mathcal{S}_n$ существует $(3k)^3$ вариантов выбрать уравнение вида $x_i = x_j x_k$, также $(3k)^3$ вариантов выбрать уравнение вида $x_i = x_j + x_k$ и только $3k$ вариантов выбрать уравнение вида $x_i = 1$. Итого для k-го уравнения есть $54k^3 + 3k$ вариантов. Отсюда получаем требуемое равенство. ■

3. Системы в форме Сколема над натуральными числами

Приведём полиномиальный генерический алгоритм для проверки разрешимости диофантовых уравнений в форме Сколема над множеством натуральных чисел $\mathbb{N}$.

Теорема 1. Проблема разрешимости диофантовых уравнений в форме Сколема над множеством $\mathbb{N}$ генерически разрешима за полиномиальное время.

Доказательство. Соответствующий генерический полиномиальный алгоритм работает на системе S следующим образом:

1) ищет в системе S уравнение вида $x_i = x_i + x_j$;
2) если такое уравнение найдётся, то система неразрешима над $\mathbb{N}$ и алгоритм выдаёт ответ 0;
3) иначе выдает ответ «?».

Покажем, что этот алгоритм дает неопределённый ответ на пренебрежимом множестве систем Сколема. Обозначим через $\mathcal{L}$ множество систем, в которых отсутствуют уравнения вида $x_i = x_i + x_j$. Число вариантов выбрать k-е уравнение в системе из множества $\mathcal{L}$ есть $54k^3 + 3k - 9k^2$. Поэтому

$$|\mathcal{L}_n| = \prod_{k=1}^{n}(54k^3 + 3k - 9k^2).$$

По лемме 1 $|\mathcal{S}_n| = \prod\limits_{k=1}^{n}(54k^3 + 3k)$, поэтому

$$\rho_n(\mathcal{L}) = \frac{|\mathcal{L}_n|}{|\mathcal{S}_n|} = \frac{\prod\limits_{k=1}^{n}(54k^3 + 3k - 9k^2)}{\prod\limits_{k=1}^{n}(54k^3 + 3k)} = \prod_{k=1}^{n}\frac{18k^3 + k - 3k^2}{18k^3 + k} = \prod_{k=1}^{n}\Big(1 - \frac{3k^2}{18k^3 + k}\Big) <$$

$$< \prod_{k=1}^{n}\Big(1 - \frac{3k^2}{18k^3 + k^3}\Big) = \prod_{k=1}^{n}\Big(1 - \frac{3}{19k}\Big) < \prod_{k=1}^{n}\Big(1 - \frac{1}{7k}\Big).$$

Для того чтобы оценить сверху последнее произведение, возведём его в степень 7:

$$\prod_{k=1}^{n}\Big(1 - \frac{1}{7k}\Big)^7 < \prod_{k=1}^{n}\Big(\big(1 - \frac{1}{7k}\big)\big(1 - \frac{1}{7k+1}\big)\ldots\big(1 - \frac{1}{7k+6}\big)\Big) = \prod_{k=7}^{n}\Big(1 - \frac{1}{k}\Big) =$$

$$= \prod_{k=7}^{n}\frac{k-1}{k} = \frac{6}{7}\cdot\frac{7}{8}\cdot\ldots\cdot\frac{n-2}{n-1}\cdot\frac{n-1}{n} = \frac{6}{n}.$$

В результате получаем

$$\rho(\mathcal{L}) = \lim_{n\to\infty}\frac{|\mathcal{L}_n|}{|\mathcal{S}_n|} \leqslant \lim_{n\to\infty}\sqrt[7]{\frac{6}{n}} = 0.$$

Теорема доказана. ■

4. Системы в форме Сколема над целыми числами

Докажем, что проблема разрешимости нормализованных систем диофантовых уравнений в форме Сколема над множеством целых чисел является неразрешимой на любом рекурсивном строго генерическом подмножестве входов.

Для произвольной нормализованной системы диофантовых уравнений в форме Сколема $S = \{S_1, \ldots, S_m\}$ рассмотрим множество систем $\mathrm{eq}(S)$, которые получаются добавлением к системе S любого количества произвольных уравнений вида $x_i = x_j x_k$ или $x_i = x_j + x_k$, где $i, j, k > 3m$, с сохранением условия нормализации. Очевидно, что любая система из $\mathrm{eq}(S)$ разрешима в целых числах тогда и только тогда, когда разрешима в целых числах система S.

Лемма 2. Для любой системы S множество $\mathrm{eq}(S)$ не является строго пренебрежимым.

Доказательство. Пусть $n > m$. Для t-го добавленного к S уравнения вида $x_i = x_j x_k$ или $x_i = x_j + x_k$, где $i, j, k > 3m$, имеется $2(3t)^3 = 54t^3$ вариантов. Поэтому

$$|\mathrm{eq}(S)_n| = \prod_{t=1}^{n-m} (54t^3).$$

Теперь по лемме 1

$$\rho_n(\mathrm{eq}(S)) = \frac{|\mathrm{eq}(S)_n|}{|\mathcal{S}_n|} = \frac{\prod\limits_{k=1}^{n-m} (54k^3)}{\prod\limits_{k=1}^{n} (54k^3 + k)} = \prod_{k=1}^{n-m} \frac{54k^2}{54k^2 + 1} \prod_{k=n-m+1}^{n} \frac{1}{54k^3 + k}.$$

Оценим снизу первое произведение:

$$\prod_{k=1}^{n-m} \frac{54k^2}{54k^2 + 1} = \prod_{k=1}^{n-m} \Big(1 - \frac{1}{54k^2 + 1}\Big) > \prod_{k=2}^{n-m+1} \Big(1 - \frac{1}{k^2}\Big) = \prod_{k=2}^{n-m+1} \frac{(k-1)(k+1)}{k^2} =$$
$$= \frac{1 \cdot 3}{2^2} \frac{2 \cdot 4}{3^2} \frac{(n-m-1)(n-m+1)}{(n-m)^2} \frac{(n-m)(n-m+2)}{(n-m+1)^2} = \frac{n-m+2}{2(n-m+1)} > \frac{1}{2}.$$

Теперь оценим второе произведение:

$$\prod_{k=n-m+1}^{n} \frac{1}{54k^3 + k} > \frac{1}{(54n^3 + n)^m}.$$

Получаем

$$\rho_n(\mathrm{eq}(S)) = \frac{|\mathrm{eq}(S)_n|}{|\mathcal{S}_n|} > \frac{1}{2(54n^3 + n)^m}.$$

Из этого неравенства следует, что последовательность $\rho_n(\mathrm{eq}(S))$ не может стремиться к нулю экспоненциально быстро. Поэтому множество $\mathrm{eq}(S)$ не является строго пренебрежимым. ■

Теорема 2. Проблема разрешимости нормализованных систем диофантовых уравнений в форме Сколема над множеством целых чисел не является строго генерически разрешимой.

Доказательство. Допустим, что существует строго генерический алгоритм $\mathcal{A}$, определяющий разрешимость диофантовых систем на некотором строго генерическом

множестве входов. Используя этот алгоритм, построим алгоритм $\mathcal{B}$, который будет определять разрешимость диофантовых систем на всём множестве входов. Тем самым получим противоречие с неразрешимостью десятой проблемы Гильберта.

Алгоритм $\mathcal{B}$ на системе S работает следующим образом: перебирает элементы $\mathrm{eq}(S)$ в порядке возрастания размера до тех пор, пока не получит систему $S' \in \mathrm{eq}(S)$, такую, что $\mathcal{A}(S') \neq ?$. Ответ $\mathcal{A}(S')$ и будет правильным ответом для исходной системы S.

То, что всегда найдётся такая система S', следует из того, что множество $\{S \in \mathcal{S} : \mathcal{A}(S) = ?\}$ строго пренебрежимо, а множество $\mathrm{eq}(S)$ не является строго пренебрежимым. ■

Заметим, что доказательство этой теоремы проходит для случая, когда решения ищутся во множестве натуральных чисел с нулём $\mathbb{N}_0$.

Автор выражает искреннюю благодарность рецензенту за полезные замечания и предложения по улучшению текста статьи.

ЛИТЕРАТУРА

1. *Матиясевич Ю. В.* Диофантовость перечислимых множеств // Доклады Академии наук СССР. 1970. Т. 191. № 2. С. 279–282.
2. *Matiyasevich Yu. and Robinson J.* Reduction of an arbitrary Diophantine equation to one in 13 unknowns // Acta Arithmetica. 1975. V. 27. P. 521–553.
3. *Jones J.* Undecidable Diophantine equations // Bull. Amer. Math. Soc. 1980. V. 3. No. 2. P. 859–862.
4. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Generic-case complexity, decision problems in group theory and random walks // J. Algebra. 2003. V. 264. No. 2. P. 665–694.
5. *Myasnikov A. and Romankov V.* Diophantine cryptography in free metabelian groups: Theoretical base // Groups, Complexity, Cryptology. 2014. V. 6. No. 2. P. 103–120.
6. *Романьков В. А.* Диофантова криптография на бесконечных группах // Прикладная дискретная математика. 2012. № 2 (16). С. 15–42.
7. *Романьков В. А.* Алгебраическая криптография. Омск: ОмГУ, 2013.
8. *Rybalov A.* Generic complexity of the Diophantine problem // Groups, Complexity, Cryptology. 2013. V. 5. No. 1. P. 25–30.
9. *Рыбалов А.* О генерической неразрешимости Десятой проблемы Гильберта // Вестник Омского университета. 2011. № 4. С. 19–22.
10. *Skolem T.* Diophantische Gleichungen. Berlin: Springer, 1938.

REFERENCES

1. *Matiyasevich Yu. V.* Diofantovost' perechislimykh mnozhestv [Diophantineity of enumerable sets]. Doklady Akademii Nauk USSR, 1970, vol. 191, no. 2, pp. 279–282. (in Russian)
2. *Matiyasevich Yu. and Robinson J.* Reduction of an arbitrary Diophantine equation to one in 13 unknowns. Acta Arithmetica, 1975, vol. 27, pp. 521–553.
3. *Jones J.* Undecidable Diophantine equations. Bull. Amer. Math. Soc., 1980, vol. 3, no. 2, pp. 859–862.
4. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Generic-case complexity, decision problems in group theory and random walks. J. Algebra, 2003, vol. 264, no. 2, pp. 665–694.
5. *Myasnikov A. and Romankov V.* Diophantine cryptography in free metabelian groups: Theoretical base. Groups, Complexity, Cryptology, 2014, vol. 6, no. 2, pp. 103–120.
6. *Roman'kov V. A.* Diofantova kriptografiya na beskonechnykh gruppakh [Diophantine cryptography over infinite groups]. Prikladnaya Diskretnaya Matematika, 2012, no. 2 (16), pp. 15–42. (in Russian)

7. *Roman'kov V. A.* Algebraicheskaya Kriptografiya [Algebraic Cryptography]. Omsk, OmSU Publ., 2013. (in Russian)
8. *Rybalov A.* Generic complexity of the Diophantine problem. Groups, Complexity, Cryptology, 2013, vol. 5, no. 1, pp. 25–30.
9. *Rybalov A.* O genericheskoy nerazreshimosti Desyatoy problemy Gil'berta [On generic undecidability of Hilbert Tenth problem]. Vestnik Omskogo Universiteta, 2011, no. 4, pp. 19–22. (in Russian)
10. *Skolem T.* Diophantische Gleichungen. Berlin, Springer, 1938.

УДК 004.8

РЕДУКЦИЯ СВЯЗЕЙ АВТОАССОЦИАТИВНОЙ ПАМЯТИ ХОПФИЛДА

М. С. Тарков

Институт физики полупроводников СО РАН, г. Новосибирск, Россия

На основе аналогии с осцилляторными сетями выполнено исследование влияния сокращения числа связей на поведение автоассоциативной сети Хопфилда. Показано, что исключение связей с весами, модули которых строго меньше максимального для данного нейрона, существенно повышает качество работы сети. При этом допустимая доля искаженных элементов входного вектора сети возрастает с увеличением его размерности.

Ключевые слова: *автоассоциативная память Хопфилда, правило Хебба, проекционный метод, осцилляторная сеть, редукция связей.*

DOI 10.17223/20710410/37/9

REDUCTION OF SYNAPSES IN THE HOPFIELD AUTOASSOCIATIVE MEMORY

M. S. Tarkov

Rzhanov Institute of Semiconductor Physics SB RAS, Novosibirsk, Russia

E-mail: tarkov@isp.nsc.ru

The auto-associative Hopfield network is a set of neurons in which the output of each neuron is the input of all other neurons, i.e. the inter-neuronal connections graph of the Hopfield network is complete. The large number of inter-neuronal connections is one of the problems of the Hopfield networks hardware implementation. A solution is the reduction (exclusion) of insignificant connections. In this paper, based on the analogy with oscillator networks, the connections number reducing effect on the auto-associative Hopfield network behavior is investigated. It is shown that the exclusion of connections with weights whose absolute values are strictly less than the maximum for a given neuron substantially improves the operation quality of the Hopfield network trained according to the Hebb's rule. As the dimension of the stored vectors increases, not only the chimeras disappear but the permissible input data noise level also increases. At the same time, the network connections number is reduced by 13–17 times. The reduction of connections in the Hopfield network, trained by the projection method, worsens its functioning quality, namely: in the network output data, there are distortions even while the reference vectors are entered. With the stored vectors dimension increasing, the allowable noise level for the reduced Hopfield — Hebb networks approaches the corresponding index for the Hopfield projection networks. Thus, given the much smaller number of connections in the reduced Hopfield — Hebb networks, these networks can successfully compete with the Hopfield projection networks for a sufficiently large stored vectors dimension.

Keywords: *auto-associative Hopfield memory, Hebb's rule, projection method, oscillatory network, reduction of connections.*

Введение

Сеть Хопфилда [1, 2] представляет собой слой нейронов с непосредственной обратной связью выхода со входом. Выходные сигналы нейронов являются одновременно входными сигналами сети: $x_i(t) = y_i(t-1)$, $t = 0, 1, 2, \ldots$, $i = 1, \ldots, N$, N — количество нейронов сети. В классической сети Хопфилда отсутствует связь выхода нейрона с собственным входом, что соответствует весам $w_{ii} = 0$, а матрица весов является симметричной: $W = W^{\mathrm{T}}$. Каждый нейрон имеет функцию активации

$$y_i = \operatorname{sign}\left(\sum_{j=1, i\neq j}^{N} w_{ij} x_j\right)$$

со значениями ± 1. Основную зависимость, определяющую сеть Хопфилда, можно представить в виде

$$y_i(t) = \operatorname{sign}\left(\sum_{j=1, i\neq j}^{N} w_{ij} y_j(t-1)\right), \ i = 1, \ldots, N, \tag{1}$$

с начальным условием $y_j(0) = x_j$.

В процессе функционирования сети Хопфилда можно выделить два режима: обучения и классификации. В режиме обучения на основе известных векторов подбираются весовые коэффициенты сети. В режиме классификации при фиксированных значениях весов и вводе конкретного начального состояния нейронов возникает переходный процесс вида (1), завершающийся в одном из локальных минимумов функции Ляпунова сети, для которого $y(t) = y(t-1)$. При вводе обучающих векторов веса вычисляются согласно обобщенному правилу Хебба

$$w_{ij} = \frac{1}{N} \sum_{k=1}^{p} x_i^k x_j^k. \tag{2}$$

Важным параметром ассоциативной памяти является её ёмкость. Под ёмкостью понимается максимальное число запомненных образов, которые классифицируются с допустимой погрешностью $\varepsilon_{\max}$. Показано [1], что при использовании для обучения правила Хебба и при $\varepsilon_{\max} = 0{,}01$ (1 % компонентов образа отличается от нормального состояния) максимальная ёмкость памяти $p_{\max} = 0{,}138N$. Такую сеть будем называть далее сетью Хопфилда — Хебба.

Проекционный метод [3, 4] обучения сети Хопфилда имеет вид итерационной зависимости матрицы весов W от последовательности обучающих векторов x^k, $k = 1, \ldots, p$:

$$y^k = (W^{k-1} - E)x^k,$$
$$W^k = W^{k-1} + \frac{y^k \cdot y^{k\mathrm{T}}}{y^{k\mathrm{T}} \cdot y^k}$$

при начальных условиях $W^0 = 0$ (E — единичная матрица). В результате предъявления p векторов матрица весов сети принимает значение $W = W^p$. Применение этого метода увеличивает максимальную ёмкость сети Хопфилда до $p_{\max} = N - 1$. Далее такую сеть будем называть проекционной.

Одной из проблем аппаратной реализации нейронных сетей является большое количество межнейронных связей (синапсов) [5 – 8]. Для автоассоциативных сетей Хопфилда эта проблема является наиболее острой в силу их полносвязности. Выходом является редукция (исключение) малозначимых связей [9]. Целью данной работы является разработка метода редукции связей и исследование влияния редукции связей на качество работы сети.

1. Редукция осцилляторной сети

Сеть из осцилляторов [10] может хранить двоичный вектор $x = (x_1, \ldots, x_N)$, $x_i \in \in \{-1, 1\}$:

$$x_i = \operatorname{sign}\left(\sum_{j=1, i\neq j}^{N} w_{ij} x_j\right),\ i = 1, \ldots, N,\ w_{ij} = \cos(\varphi_i(t) - \varphi_j(t)),$$

где $\varphi_i(t)$ — фаза i-го осциллятора.

При использовании правила Хебба (2) сеть осцилляторов хранит большое количество ложных образов (химер). С целью их исключения в [10] предложен способ модификации весов (1):

$$w_{ij}^{'} = \begin{cases} +1, w_{ij} = +1, \\ -1, w_{ij} = -1, \\ 0, -1 < w_{ij} < +1. \end{cases} \tag{3}$$

Иными словами, исключаются связи с весами, модуль которых меньше 1, поскольку такие связи делают возможным появление химер [1].

2. Редукция сети Хопфилда

Правило (3) редукции весов можно перенести на произвольную сеть Хопфилда. Рассмотрим произвольную строку $w_i = (w_{i1}, \ldots, w_{iN})$, $i \in \{1, \ldots, N\}$, матрицы весов этой сети (вектор весов нейрона). Исходя из вида ступенчатой функции активации sign, можно утверждать, что для этой функции важны лишь относительные величины элементов w_{ij} в строке w_i. Поэтому, не меняя результатов работы сети, можно нормировать величины w_{ij} следующим образом:

$$w_{ij} \leftarrow \frac{w_{ij}}{\max\limits_{j=1,\ldots,N} |w_{ij}|}.$$

Нормированные величины удовлетворяют неравенству $-1 \leqslant w_{ij} \leqslant +1$. Тогда можно ввести величину сдвига фазы $\varphi_{ij} = \arccos(w_{ij})$ между компонентой $x_j(t) = \pm 1$ вектора входных сигналов i-го нейрона и его выходным сигналом $x_i(t+1) = \operatorname{sign}\left(\sum\limits_{j=1, i\neq j}^{N} w_{ij} x_j(t)\right)$, то есть

$$x_i(t+1) = \operatorname{sign}\left(\sum_{j=1, i\neq j}^{N} \cos\varphi_{ij} x_j(t)\right),$$

и, согласно [10], к коэффициентам $w_{ij} = \cos\varphi_{ij}$ можно применить преобразование (3), то есть исключить из сети Хопфилда связи, имеющие веса $-1 < w_{ij} < +1$.

С целью сохранения симметрии матрицы весов правило редукции (3) для сетей Хопфилда приобретает следующий вид: вес w_{ij} обнуляется одновременно с весом w_{ji} при одновременном выполнении условий $|w_{ij}| < |w_{i,\max}|$ и $|w_{ji}| < |w_{j,\max}|$, где $w_{i,\max}$ и $w_{j,\max}$ — максимальные веса в строках i и j соответственно.

3. Эксперименты

В экспериментах в качестве эталонных объектов использовались биполярные векторы, полученные кодированием изображений десятичных цифр (рис. 1) (черным пикселям ставится в соответствие -1, белым $+1$). Эксперименты проводились для изображений с размерами 8×8, 16×16, 32×32, 64×64 и 128×128 пикселей.

Рис. 1. Эталонные изображения

В табл. 1 представлены количества связей в сети Хопфилда до (E_{before}) и после (E_{after}) редукции, а также отношение этих величин (коэффициент сокращения числа связей). Из табл. 1 следует, что предложенный метод позволяет сократить количество связей в сети Хопфилда — Хебба более чем на порядок.

Таблица 1

Количество связей в сети Хопфилда

$n \times n$	8×8	16×16	32×32	64×64	128×128
E_{before}	4032	65280	1047552	16773120	268419072
E_{after}	297	3840	64512	1044480	16744448
$E_{\text{after}}/E_{\text{before}}$	0,074	0,059	0,062	0,062	0,062

Установлено, что классическая сеть Хопфилда, обученная по правилу Хебба, не содержит химер (рис. 2), если количество хранимых эталонов не более 3. При выполнении редукции связей в этой сети для всех изображений цифр размером 8×8 работа сети существенно улучшается, хотя не все изображения цифр восстанавливаются полностью даже при нулевом уровне шумов (искажений пикселей) (рис. 3).

Рис. 2. Химеры, порождаемые сетью Хопфилда, обученной по правилу Хебба, при числе эталонов равном 10

Рис. 3. Восстановленные изображения цифр размером 8×8 после редукции связей в сети Хопфилда — Хебба

Будем называть уровень шума (рис. 4) допустимым, если для заданного набора эталонных изображений сеть Хопфилда полностью фильтрует шум. При увеличении размеров изображений сеть Хопфилда — Хебба с редуцированными связями не только не порождает химер, но и позволяет повысить максимально допустимый уровень шума ξ_{Hebb} входных данных (табл. 2).

Попытка редукции связей в проекционной сети Хопфилда положительного результата не дала. В этом случае для эталонных изображений всех вышеуказанных размеров редуцированная сеть Хопфилда порождала искажённые изображения даже при

Рис. 4. Пример зашумленного изображения цифр при уровне шума 57 %

отсутствии искажений во входных данных. Проведено исследование нередуцированной проекционной сети на устойчивость к шумам во входных данных. Результаты представлены в табл. 2.

Таблица 2

Допустимый уровень шума ξ в процентах

$n \times n$	8×8	16×16	32×32	64×64	128×128
$\xi_{\text{Projection}}$	14	33	48	64	89
ξ_{Hebb}	–	3	22	57	88

На рис. 5 проводится сравнение редуцированной сети Хопфилда — Хебба и проекционной нередуцированной сети Хопфилда по допустимому проценту искаженных пикселей во входных данных. Из этого сравнения следует, что с увеличением размера хранимых изображений (числа нейронов сети Хопфилда) устойчивость к шумам ξ_{Hebb} редуцированной сети Хопфилда — Хебба приближается к соответствующему показателю $\xi_{\text{Projection}}$ проекционной сети Хопфилда. При этом количество связей в редуцированной сети более чем на порядок меньше числа связей классической сети Хопфилда (табл. 1).

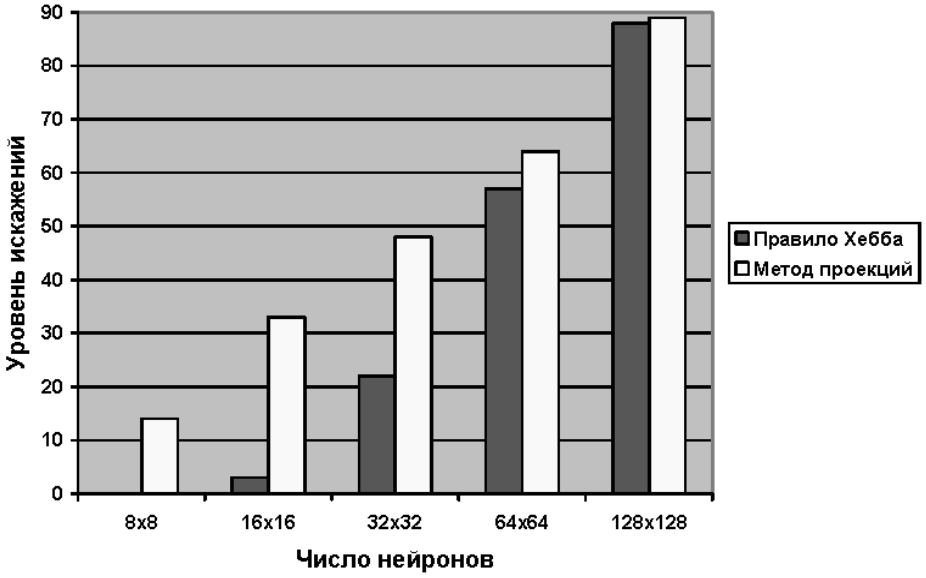

Рис. 5. Допустимый уровень процента искаженных пикселей

Заключение

Выполнено исследование влияние редукции связей сети Хопфилда на качество её функционирования. Показано, что:

1. Редукция связей в сети Хопфилда — Хебба существенно улучшает качество её функционирования. С ростом размерности хранимых векторов не только исчезают химеры, но и возрастает допустимый уровень шума во входных данных. При этом для изображений размерами 8×8, 16×16, 32×32, 64×64 и 128×128 пикселей количество связей сети сокращается в 13–17 раз, то есть более чем на порядок.

2. Редукция связей в проекционной сети Хопфилда ухудшает качество её функционирования: в выходных данных сети появляются искажения даже при вводе эталонных векторов.

3. С ростом размерности хранимых векторов допустимый уровень шума для редуцированных сетей Хопфилда — Хебба приближается к соответствующему показателю для классических проекционных сетей Хопфилда.

Таким образом, с учётом гораздо меньшего числа связей в редуцированных сетях Хопфилда — Хебба эти сети при достаточно большой размерности хранимых векторов могут успешно конкурировать с классическими проекционными сетями Хопфилда.

ЛИТЕРАТУРА

1. *Осовский С.* Нейронные сети для обработки информации. М.: Финансы и статистика, 2002. 344 с.
2. *Hopfield J.* Neural networks and physical systems with emergent collective computational abilities // Proc. National Academy of Science USA. 1982. V. 79. P. 2554–2558.
3. *Personnaz L., Guyon I., and Dreyfus G.* Collective computational properties of neural networks: New learning mechanisms // Phys. Rev. A. 1986. V. 34. No. 5. P. 4217–4228.
4. *Michel A. N. and Liu D.* Qualitative Analysis and Synthesis of Recurrent Neural Networks. N. Y.: Marcel Dekker Inc., 2002. 504 p.
5. *Hu S. G., Liu Y., Liu Z., et al.* Associative memory realized by a reconfigurable memristive Hopfield neural network // Nature Commun. 2015. V. 6. Article no. 7522.
6. *Тарков М. С.* Сеть Хопфилда с межнейронными соединениями на основе мемристорных мостов // Труды XVIII Междунар. науч.-технич. конф. «НЕЙРОИНФОРМАТИКА-2016». М.: МИФИ, 2016. Ч. 3. С. 91–100.
7. *Tarkov M. S.* Hopfield network with interneuronal connections based on memristor bridges // LNCS. 2016. V. 9719. P. 196–203.
8. *Tarkov M. S.* Oscillatory neural associative memories with synapses based on memristor bridges // Optical Memory and Neural Networks (Information Optics). 2016. V. 25. No. 4. P. 219–227.
9. *Горбань А. Н., Россиев Д. А.* Нейронные сети на персональном компьютере. Новосибирск: Наука, 1996. 276 с.
10. *Maffezzoni P., Bahr B., Zhang Z., and Daniel L.* Analysis and design of Boolean associative memories made of resonant oscillator arrays // IEEE Trans. Circuits Systems I: Regular Papers. 2016. V. 63. No. 8. P. 1964–1973.

REFERENCES

1. *Osovskiy S.* Neironnye seti dlya obrabotki informatsii [Neural networks for data processing]. Moscow, Finansy i Statistika, 2002. 344 p. (in Russian)
2. *Hopfield J.* Neural networks and physical systems with emergent collective computational abilities. Proc. National Academy of Science USA, 1982, vol. 79, pp. 2554–2558.
3. *Personnaz L., Guyon I., and Dreyfus G.* Collective computational properties of neural networks: New learning mechanisms. Phys. Rev. A, 1986, vol. 34, no. 5, pp. 4217–4228.

4. *Michel A. N. and Liu D.* Qualitative Analysis and Synthesis of Recurrent Neural Networks. N. Y., Marcel Dekker Inc., 2002. 504 p.
5. *Hu S. G., Liu Y., Liu Z., et al.* Associative memory realized by a reconfigurable memristive Hopfield neural network. Nature Commun., 2015, vol. 6, article no. 7522.
6. *Tarkov M. S.* Set Hopfilda c mezhneyronnymi soedineniyami na osnove memristornykh mostov [Hopfield Network with interneuronal connections on the base of memristor bridges]. Proc. NEIROINFORMATIKA-2016, Moscow, MEPhI Publ., 2016, part 3, pp. 91–100. (in Russian)
7. *Tarkov M. S.* Hopfield network with interneuronal connections based on memristor bridges. LNCS, 2016, vol. 9719, pp. 196–203.
8. *Tarkov M. S.* Oscillatory neural associative memories with synapses based on memristor bridges. Optical Memory and Neural Networks (Information Optics), 2016, vol. 25, no. 4, pp. 219–227.
9. *Gorban A. N. and Rossiev D. A.* Neuronnye seti na personalnom kompyutere [Neural networks on personal computer]. Novosibirsk, Nauka Publ., 1996. 276 p.
10. *Maffezzoni P., Bahr B., Zhang Z., and Daniel L.* Analysis and design of Boolean associative memories made of resonant oscillator arrays. IEEE Trans. Circuits Systems I: Regular Papers, 2016, vol. 63, no. 8, pp. 1964–1973.

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ В ДИСКРЕТНОЙ МАТЕМАТИКЕ

УДК 519.7

ОПТИМАЛЬНАЯ МАРШРУТИЗАЦИЯ ПО ОРИЕНТИРАМ В НЕСТАЦИОНАРНЫХ СЕТЯХ

В. В. Быкова, А. А. Солдатенко

Сибирский федеральный университет, г. Красноярск, Россия

Рассмотрена задача Time-Dependent Shortest-Path (TDSP), которая является расширением задачи о кратчайшем пути в графе. Сеть представляется ориентированным графом $G = (V, E)$, в котором для каждой дуги $(x, y) \in E$ определены две функции: $w_{xy}(t)$ — время, необходимое для передвижения по дуге (x, y), и $F_{xy}(t)$ — время прибытия в вершину y при условии, что старт из вершины x осуществлён в момент времени t. Такую сеть называют нестационарной, а наименьшее время передвижения из стартовой вершины в целевую интерпретируют как оптимальный маршрут или кратчайший путь между этими вершинами. В работе задача TDSP исследована для полиномиально разрешимого случая, когда функции прибытия являются монотонными. Двухфазный алгоритм ALT (A* with Landmarks & Triangle) — один из современных алгоритмов, способных быстро решать задачу TDSP на графах большой размерности. В работе определено и доказано достаточное условие корректности алгоритма ALT для задачи TDSP: сеть должна отвечать неравенству треугольника, заданного для промежутков времени передвижения по узлам сети. Особенность предложенного неравенства треугольника — его определение через «оптимистичные» веса дуг, когда возможно беспрепятственное движение по дугам. Показано, что это неравенство треугольника верно всегда, если веса дуг заданы отношениями длин дуг к скорости передвижения по ним и справедливо неравенство треугольника для расстояний между узлами сети.

Ключевые слова: *нестационарные сети, оптимальная маршрутизация, корректность алгоритма ALT.*

DOI 10.17223/20710410/37/10

OPTIMAL ROUTING BY LANDMARKS IN THE TIME-DEPENDENT NETWORKS

V. V. Bykova, A. A. Soldatenko

Siberian Federal University, Krasnoyarsk, Russia

E-mail: bykvalen@mail.ru, glinckon@gmail.com

The Time-Dependent Shortest-Path problem (TDSP) is considered. This is an extension of the shortest path problem in a graph. TDSP problem arises in designing and operating telecommunication and transport networks. In such a network, we need to consider the time and the possibility of appearance of predictable situations, for example, traffic jams or traffic reduction. In this case, the network is represented with

a digraph $G = (V, E)$ in which, for each arc $(x, y) \in E$, two functions are defined. The first one, $w_{xy}(t)$, is the time required for move along an arc (x, y). The second function, $F_{xy}(t)$, is the time of arrival at the vertex y if the movement starts from the vertex x at the time t. Such a graph is called a time-dependent network and the minimum time for movement from vertex x to vertex y is interpreted as the optimal route or the shortest path between these vertices. In this paper, TDSP problem is studied for polynomial case when the arrival function is monotonous. The two-phased ALT (A* with Landmarks & Triangle) algorithm is one of the modern algorithms which are able to fast solve the TDSP problem for the large dimension graphs. There exist many experimental works done to improve ALT algorithm. However, the theoretical foundation of the ALT algorithm applicability to different classes of graphs is desperately needed. In this paper, we establish and prove sufficient conditions for correctness of ALT algorithm with respect to TDSP problem. These conditions define requirements for a time-dependent network such that if the requirements are fulfilled, then the ALT algorithm finds the exact solution of TDSP problem — a time-dependent shortest path in the network. According to the conditions, the network must satisfy triangle inequality for the time intervals of moving between the network nodes. A distinguishing feature of this concept of the triangle inequality is its definition through so called optimistic arc weights providing an unimpeded movement along the arcs. We show that in the network, where a weight of an arc is defined as ratio of its length to velocity of moving on it, this triangle inequality is always true if the triangle inequality for the distance between the network nodes is true.

Keywords: *time-dependent network, optimal routing, correctness of ALT algorithm.*

Введение

Задача поиска кратчайшего пути в графе (Shortest-Paths, SP) — хорошо известная задача комбинаторной оптимизации, имеющая многие реальные приложения. Задача SP состоит в нахождении кратчайшего пути между вершинами s и d (стартовой и целевой соответственно) в заданном графе $G = (V, E)$ [1, 2]. При проектировании и эксплуатации современных телекоммуникационных и транспортных сетей часто приходится иметь дело с расширением задачи SP, когда требуется учитывать временной фактор и возможность возникновения в отдельные промежутки времени таких предсказуемых ситуаций, как снижение объёма трафика и наличие пробок в сети. В этом случае сеть представляется ориентированным графом $G = (V, E)$, в котором для каждой дуги $(x, y) \in E$ определены две функции: $w_{xy}(t)$ — время, необходимое для передвижения по дуге (x, y), и $F_{xy}(t)$ — время прибытия в вершину y при условии, что старт из вершины x осуществлён в момент времени t. Такую сеть называют нестационарной, а наименьшее время передвижения из стартовой вершины в целевую интерпретируют как оптимальный маршрут или кратчайший путь между этими вершинами, причём время передвижения по этому пути всегда зависит от момента выхода из стартовой вершины. В литературе задача поиска кратчайшего пути в нестационарной сети носит название Time-Dependent Shortest-Path problem (TDSP) [3, 4]. Известно, что TDSP для нестационарной сети общего вида без каких-либо ограничений на топологию сети и функции прибытия является NP-трудной задачей [3]. В случае, когда функции прибытия являются монотонными, задача TDSP полиномиально разрешима [4, 5]. В данной работе рассматривается именно этот полиномиально разрешимый случай.

К настоящему времени уже предложено много алгоритмов решения задачи TDSP. Самый известный из них — алгоритм Дейкстры, который при условии неотрицательно-

сти значений весов дуг находит точное решение задач SP и TDSP в графе $G = (V, E)$ за время $O(|V|^2)$ [2, 5]. Современные телекоммуникационные и транспортные сети могут быть огромными — содержать миллионы узлов, и требуется найти оптимальный маршрут из стартовой вершины в целевую за доли секунды. Экспериментально установлено, что в подобных условиях алгоритм Дейкстры работает неприемлемо долго [6, 7]. Чтобы справиться с этой проблемой, предложены различные методы ускорения алгоритма Дейкстры, позволяющие вычислять оптимальный маршрут за несколько микросекунд даже в огромных сетях [6 – 10].

Большинство этих методов основано на двухфазном подходе. Первая фаза — предобработка графа, моделирующего исходную сеть. На второй фазе с помощью алгоритма Дейкстры (или какой-либо его версии) осуществляется поиск оптимального маршрута в графе, полученном после предобработки. Предобработка сводится к просмотру исходного графа и анализу его структуры с целью извлечения информации, позволяющей ускорить вторую фазу алгоритма. Двухфазные алгоритмы принято разделять на следующие классы:

— иерархические алгоритмы, которые основаны на многоуровневом представлении исходного графа с выполнением алгоритма Дейкстры для каждого выделенного уровня [8];
— алгоритмы маркировки, в которых исходный граф разбивается на несколько примерно равных частей и маршрутизация осуществляется по меткам дуг, находящихся на стыке этих частей [9];
— алгоритмы маршрутизации по ориентирам, в которых некоторое множество вершин исходного графа рассматривается в роли ориентиров и используется для целенаправленного поиска оптимального маршрута [7, 10].

Большинство двухфазных алгоритмов первоначально разрабатывались для задачи SP и поэтому лишь немногие из них пригодны для решения задачи TDSP и учёта динамических сценариев, возникающих в реальных сетях и изменяющих их структуру. По мнению многих авторов, алгоритм ALT (A* with Landmarks & Triangle), являющийся современным представителем класса алгоритмов маршрутизации по ориентирам, хорошо подходит для TDSP и оптимальной маршрутизации в динамических графах [10 – 12]. В пользу алгоритма ALT приводятся следующие аргументы. Алгоритм ALT на этапе предобработки расставляет определенное число ориентиров в вершинах графа, не трансформируя его, т. е. не изменяя состава вершин и дуг графа, весов дуг и функций прибытия. Это позволяет при изменении графа не повторять полностью фазу предобработки, а лишь вносить локальные изменения в данные, используемые для ускорения второй фазы алгоритма ALT. На второй фазе ALT для поиска оптимального маршрута используется алгоритм A* — версия алгоритма Дейкстры, работающая с потенциальными функциями, вычисленными на основе ориентиров [13]. Использование в A* потенциальных функций позволяет на практике находить оптимальный маршрут значительно быстрее алгоритма Дейкстры. Например, для европейской дорожной сети, состоящей из более чем 1,5 миллионов вершин, алгоритм ALT в среднем работает в 20 раз быстрее алгоритма Дейкстры [10, 14]. Алгоритм ALT был предложен в 2005 г. А. Гольдбергом и К. Харрельсоном [7]. Сочетание алгоритма ALT с другими методами ускорения, а также совершенствование процедуры расстановки ориентиров — современные направления развития данного алгоритма [6, 10, 11]. Однако алгоритм ALT применим не ко всем графам, поскольку для корректной работы алгоритма A* необходимо, чтобы потенциальные функции, вычисленные на основе заданных весов дуг

исходного графа, обладали свойствами допустимости и преемственности [13]. В данном случае под корректностью алгоритма понимается его способность находить точное решение поставленной оптимизационной задачи [2].

Существует достаточно много экспериментальных работ, посвящённых совершенствованию алгоритма ALT. Однако актуальны теоретические основы применимости данного алгоритма к различным классам графов, в частности к нестационарным сетям. Практически отсутствуют работы, направленные на доказательство корректности алгоритма ALT при различных способах задания весов дуг и функций прибытия в задаче TDSP. Основным результатом данной работы является определение достаточных условий корректности алгоритма ALT для задачи TDSP и их доказательство.

1. Формулировка задачи TDSP

Введём обозначения, необходимые для формулировки задачи TDSP. Понятия и обозначения теории графов, которые не вводятся ниже, общеприняты и взяты из [1].

Пусть задан ориентированный граф $G = (V, E)$ без кратных дуг и петель (далее просто граф), в котором для всякой дуги $(x, y) \in E$ определена весовая функция

$$w_{xy}(t) = \frac{l_{xy}}{v_{xy}(t)} \geqslant 0. \tag{1}$$

В формуле (1) величину l_{xy} интерпретируем как длину дуги (x, y), $v_{xy}(t)$ — скорость движения по дуге (x, y), а $w_{xy}(t)$ — время передвижения из вершины x в вершину y при условии, что старт из вершины x осуществлён в момент времени t. Считаем, что $v_{xy}(t) > 0$, $l_{xy} \geqslant 0$, а также что расстояния между вершинами графа G являются постоянными величинами и подчиняются неравенству треугольника. Кроме того, полагаем, что величина $t \geqslant 0$ измеряется в условных единицах времени и принимает значения из некоторого конечного множества T. Таким образом, $w_{xy}(t)$ и $v_{xy}(t)$ — дискретные функции с конечным множеством значений. В данном случае расстояние между вершинами графа — кратчайший путь между этими вершинами в обычном для теории графов понимании [1].

Сопоставим дуге $(x, y) \in E$ функцию прибытия

$$F_{xy}(t) = t + w_{xy}(t),$$

где t — время отправления из вершины x; $F_{xy}(t)$ — время прибытия в вершину y при движении по дуге (x, y). Поскольку $t \geqslant 0$, $w_{xy}(t) \geqslant 0$, то всегда

$$F_{xy}(t) \geqslant t \geqslant 0. \tag{2}$$

Неравенство (2) отражает непосредственный ход времени: «отправляясь из вершины x в момент времени t, невозможно прибыть в вершину y раньше времени t».

Если для любых моментов времени $0 \leqslant t_1 \leqslant t_2$ верно

$$F_{xy}(t_1) \leqslant F_{xy}(t_2),$$

то говорят, что функция прибытия дуги (x, y) монотонная. Граф $G = (V, E)$, отвечающий всем указанным выше предположениям, включая монотонность функций прибытия всех его дуг, принято называть нестационарной сетью, удовлетворяющей условию First-In First-Out (FIFO) [3, 4].

Последовательность вершин $P = (x_0, x_1, \ldots, x_k)$ графа $G = (V, E)$, в которой $s = x_0$, $d = x_k$, $(x_i, x_{i+1}) \in E$, $i = 0, 1, \ldots, k-1$, задаёт путь из вершины s в вершину d. Если

при этом начало движения осуществлено в момент времени t_s, то данный путь будем обозначать (s, d, t_s) или (P, t_s). Существование в графе (s, d, t_s)-пути указывает, что вершина d достижима из вершины s. В этом случаем будем записывать $s \underset{P}{\rightsquigarrow} d$. Вес (P, t_s)-пути равен, по определению,

$$w(P, t_s) = t_s + \sum_{i=0}^{k-1} w_{x_i x_{i+1}}(t_i), \tag{3}$$

где $t_0 = t_s$; $t_{i+1} = F_{x_i x_{i+1}}(t_i)$, $i = 0, \ldots, k-2$, а вес кратчайшего (s, d, t_s)-пути

$$\mathrm{dist}(s, d, t_s) = \min_P \{w(P, t_s) \colon s \underset{P}{\rightsquigarrow} d\}. \tag{4}$$

Заметим, что значения величин $w(P, t_s)$ и $\mathrm{dist}(s, d, t_s)$ всегда могут быть вычислены по формулам (3) и (4), если вершина d достижима из вершины s. Величину $w(P, t_s)$ будем трактовать как время прибытия в вершину d при прохождении (s, d, t_s)-пути, а величину $\mathrm{dist}(s, d, t_s)$ — как самое раннее время прибытия в d при условии, что отправление из вершины s осуществлено в момент времени t_s. Тройку величин (s, d, t_s), где s и d — стартовая и целевая вершины соответственно, t_s — стартовое время, будем называть запросом на поиск в нестационарной сети $G = (V, E)$ кратчайшего (s, d, t_s)-пути, или кратко (s, d, t_s)-запросом.

С использованием введённых понятий и обозначений задача TDSP формулируется следующим образом.

Time-Dependent Shortest-Path problem
Заданы: нестационарная сеть $G = (V, E)$ с условием FIFO, (s, d, t_s)-запрос.
Требуется: найти значение $\mathrm{dist}(s, d, t_s)$ и последовательность вершин, образующих кратчайший (s, d, t_s)-путь.

Очевидно, что в указанной постановке задача TDSP всегда имеет решение, если вершина d достижима из вершины s. Точное решение может быть найдено за время $\mathrm{O}(|V|^2)$ с помощью алгоритма Дейкстры. Корректность алгоритма Дейкстры гарантируется условием FIFO для сети G и неотрицательностью значений весов всех её дуг. При применении других алгоритмов для нахождения точного решения задачи TDSP могут возникать дополнительные требования к G, например, для алгоритма ALT необходимо, чтобы сеть G удовлетворяла неравенству треугольника.

2. Определение неравенства треугольника и потенциальных функций для нестационарной сети

Алгоритм ALT осуществляет целенаправленный поиск оптимального маршрута по ориентирам. Под ориентиром понимается некоторая выделенная вершина исходного графа. Алгоритм ALT состоит из двух фаз: на первой фазе выполняется расстановка ориентиров и вычисление на их основе потенциальных функций, а на второй фазе — поиск оптимального маршрута с помощью алгоритма A*. Потенциальные функции предназначены для оценки снизу значения кратчайшего пути от текущей до целевой вершины.

Как отмечено выше, алгоритм A* находит точное решение задачи SP для исходного графа $G = (V, E)$ за время $\mathrm{O}(|V|^2)$, если потенциальные функции обладают свойствами допустимости и преемственности. Для задачи SP данные свойства всегда выполнимы, если расстояния между вершинами графа G подчиняются неравенству треугольника [13].

Определим неравенство треугольника и потенциальные функции для нестационарной сети. Для этого введём величину максимальной скорости, с которой можно двигаться по всякой дуге исходного графа:

$$v_{\max} = \max_{(x,y)\in E}\{\max_{t\in T}[v_{xy}(t)]\} > 0.$$

Заметим, что величина $v_{\max}$ постоянна для $G = (V, E)$ на рассматриваемом промежутке времени T. Вместо формулы (1) для каждой дуги $(x, y) \in E$ графа G определим её «оптимистичный» вес следующим образом:

$$w'_{xy} = \frac{l_{xy}}{v_{\max}}.$$

Значение w'_{xy} не зависит от t и равно времени передвижения от вершины x к вершине y при условии отсутствия каких-либо препятствий, приводящих к снижению скорости по этой дуге. Другими словами, это минимальное время передвижения по дуге (x, y). Отсюда для всякой дуги $(x, y) \in E$ и любого $t \in T$ верна оценка

$$w'_{xy} \leqslant w_{xy}(t).$$

Сопоставим пути $P = (x_0, x_1, \ldots, x_k)$, где $s = x_0$, $d = x_k$, $(x_i, x_{i+1}) \in E$, $i = 0, 1, \ldots, k-1$, графа $G = (V, E)$ следующую величину:

$$w'(P) = \sum_{i=0}^{k-1} w'_{x_i x_{i+1}}. \tag{5}$$

Данная величина не зависит от t_s и равна минимальному времени передвижения по пути P из вершины s в вершину d. Очевидна оценка $w'(P) \leqslant w(P, t_s)$.

Аналогично для кратчайшего (s, d, t_s)-пути имеем

$$\operatorname{dist}'(s, d) = \min_P\{w'(P)\colon s \overset{P}{\rightsquigarrow} d\}; \tag{6}$$

$$\operatorname{dist}'(s, d) \leqslant \operatorname{dist}(s, d, t_s). \tag{7}$$

Для заданных s, d значение $\operatorname{dist}'(s, d)$ — наименьшее время, необходимое для передвижения в графе G из вершины s в вершину d с использованием всех возможных путей.

Определим неравенство треугольника для нестационарной сети $G = (V, E)$ через величины, вычисляемые формулой (6):

$$\operatorname{dist}'(x, y) + \operatorname{dist}'(y, z) \geqslant \operatorname{dist}'(x, z), \tag{8}$$

где $x, y, z \in V$. Данное неравенство означает, что наименьшее время передвижения от вершины x в вершину z по кратчайшему пути (в смысле формулы (6)) всегда не больше времени движения в обход этого пути через некоторую вершину y. Если для любых вершин $x, y, z \in V$ нестационарной сети $G = (V, E)$ верно неравенство (8), то такую сеть назовём нестационарной метрической сетью.

Пусть $L \subseteq V$ — множество вершин нестационарной метрической сети $G = (V, E)$, в которых установлены ориентиры. Для каждой вершины $x \in V$ и всякого ориентира $l \in L$ определим следующие функции:

$$\pi_{l+}(x) = \operatorname{dist}'(l, d) - \operatorname{dist}'(l, x); \tag{9}$$

$$\pi_{l-}(x) = \operatorname{dist}'(x, l) - \operatorname{dist}'(d, l); \tag{10}$$

$$\pi_l(x) = \max\{0, \pi_{l+}(x), \pi_{l-}(x)\}. \tag{11}$$

Неотрицательную величину

$$\pi_L(x) = \max_{l \in L} \pi_l(x) \tag{12}$$

назовём потенциальной функцией вершины x относительно множества ориентиров L. Переход в (9)–(11) от длин кратчайших путей, вычисляемых по формуле (4), к их нижним оценкам (6) — особенность предлагаемого в данной работе определения потенциальных функций для нестационарной сети. В остальном это определение соответствует определению потенциальных функций, введённому в [7].

3. Достаточное условие корректности алгоритма ALT для нестационарной метрической сети

Опираясь на работы [7, 13], определим свойства допустимости и преемственности для потенциальных функций применительно к нестационарным сетям. Потенциальная функция, удовлетворяющая условию

$$0 \leqslant \pi_L(x) \leqslant \operatorname{dist}(x, d, t_x), \tag{13}$$

называется допустимой. Потенциальная функция называется преемственной, если

$$\pi_L(s) \leqslant \operatorname{dist}(s, x, t_s) + \pi_L(x). \tag{14}$$

Для корректной работы алгоритма A* требуется выполнимость неравенств (13) и (14) для любых $s, x, d \in V$, $t_s \in T$ и произвольного непустого множества ориентиров $L \subseteq V$.

Теорема 1. Для нестационарной метрической сети $G = (V, E)$ и непустого множества ориентиров $L \subseteq V$ неравенство треугольника (8) задаёт достаточное условие для допустимости и преемственности потенциальных функций, определённых формулами (9)–(12).

Доказательство. Докажем допустимость потенциальных функций. Исходя из (7) и (8), для произвольных $x, d \in V$, $l \in L$ и $t_x \in T$ имеем

$$\pi_{l+}(x) = \operatorname{dist}'(l, d) - \operatorname{dist}'(l, x) \leqslant \operatorname{dist}'(x, d) \leqslant \operatorname{dist}(x, d, t_x),$$
$$\pi_{l-}(x) = \operatorname{dist}'(x, l) - \operatorname{dist}'(d, l) \leqslant \operatorname{dist}'(x, d) \leqslant \operatorname{dist}(x, d, t_x).$$

Поскольку всегда $0 \leqslant \operatorname{dist}(x, d, t_x)$, то также

$$\pi_l(x) = \max\{0, \pi_{l+}(x), \pi_{l-}(x)\} \leqslant \operatorname{dist}(x, d, t_x),$$
$$\pi_L(x) = \max_{l \in L} \pi_l(x) \leqslant \operatorname{dist}(x, d, t_x),$$

т. е. условие (13) справедливо для любых $x, d \in V$, $t_x \in T$ и непустого множества ориентиров $L \subseteq V$.

Докажем преемственность потенциальных функций. Согласно (7) и (8), для всякого ориентира $l \in L$ и произвольных $s, x \in V$, $t_s \in T$ верны соотношения

$$\operatorname{dist}'(s, x) \leqslant \operatorname{dist}(s, x, t_s),$$
$$\operatorname{dist}'(l, x) - \operatorname{dist}'(l, s) \leqslant \operatorname{dist}'(s, x).$$

Отсюда получаем

$$\operatorname{dist}'(l, x) - \operatorname{dist}'(l, s) \leqslant \operatorname{dist}(s, x, t_s). \tag{15}$$

Добавим в обе части неравенства (15) положительную величину $\mathrm{dist}'(l,d)$:

$$\mathrm{dist}'(l,d)+\mathrm{dist}'(l,x)-\mathrm{dist}'(l,s)\leqslant \mathrm{dist}(s,x,t_s)+\mathrm{dist}'(l,d).$$

Перенесём $\mathrm{dist}'(l,x)$ в правую часть неравенства:

$$\mathrm{dist}'(l,d)-\mathrm{dist}'(l,s)\leqslant \mathrm{dist}(s,x,t_s)+\mathrm{dist}'(l,d)-\mathrm{dist}'(l,x).$$

С учётом (9) данное неравенство принимает вид

$$\pi_{l+}(s)\leqslant \mathrm{dist}(s,x,t_s)+\pi_{l+}(x).$$

Аналогичным образом для ориентира $l\in L$ и формулы (10) получаем

$$\pi_{l-}(s)\leqslant \mathrm{dist}(s,x,t_s)+\pi_{l-}(x).$$

Тогда, согласно (11), для ориентира $l\in L$ при любых $s,x,d\in V$ и $t_s\in T$ верно неравенство

$$\pi_{l}(s)\leqslant \mathrm{dist}(s,x,t_s)+\pi_{l}(x).$$

Значит, условие (14) всегда справедливо для любых $s,x,d\in V$, $t_s\in T$ и непустого множества ориентиров $L\subseteq V$. ■

Следствие 1. Если веса дуг нестационарной сети $G=(V,E)$ определены формулой (1) и верно предположение о выполнимости неравенства треугольника для расстояний между вершинами графа G, то неравенство (8) всегда справедливо.

В самом деле, доказательство теоремы 1 базируется на неравенствах (7) и (8). Когда веса дуг нестационарной сети G заданы формулой (1), равенство (5) принимает следующий вид:

$$w'(P)=\frac{1}{v_{\max}}\sum_{i=0}^{k-1} l_{x_ix_{i+1}}.$$

Тогда в (6) величина $\mathrm{dist}'(s,d)$ является расстоянием между вершинами s и d, вычисленным на основе длин дуг графа G, неравенство (7) остаётся верным, а неравенство (8) вырождается в неравенство треугольника для расстояний между вершинами этого графа.

Заметим, что предположение о выполнимости неравенства треугольника для расстояний является естественным свойством графов, моделирующих широкий класс телекоммуникационных и транспортных сетей. Согласно следствию 1, для таких сетей достаточное условие теоремы 1 не требует специальной проверки, поскольку всегда справедливо. Этот факт чрезвычайно важен для сетей большой размерности, так как подобная проверка может быть трудоёмкой и сопоставимой по времени работы с алгоритмом Дейкстры. Требование выполнимости неравенства треугольника для корректной работы алгоритма ALT при решении задачи SP было ранее определено и доказано в [7, 13]. Теорема 1 определяет подобное требование для задачи TDSP.

Предложенные в п. 2 формулы (9)–(12) описывают процесс нахождения потенциальных функций для нестационарной сети. По трудоёмкости этот процесс сопоставим с вычислением потенциальных функций для задачи SP. Однако потенциальные функции, определённые формулами (9)–(12), не всегда являются наилучшими в смысле нижних оценок (13), что влечёт снижение быстродействия алгоритма A*. Возможные направления улучшения нижних оценок — совершенствование стратегии расстановки ориентиров и нахождение новых способов определения допустимых и преемственных потенциальных функций для нестационарной сети.

ЛИТЕРАТУРА

1. *Емеличев В. А., Мельников О. И., Сарванов В. И., Тышкевич Р. И.* Лекции по теории графов. М.: Книжный дом «Либроком», 2012. 390 с.
2. *Кормен Т. Х., Лейзерсон Ч. И., Ривест Р. Л., Штайн К.* Алгоритмы. Построение и анализ. М.: Вильямс, 2016. 1328 с.
3. *Sherali H. D., Ozbay K., and Subramanian S.* The time-dependent shortest pair of disjoint paths problem: Complexity, models, and algorithms // J. Networks. 1998. V. 31. No. 4. P. 259–272.
4. *Kaufman D. E. and Smith R. L.* Fastest paths in time-dependent networks for intelligent vehicle-highway systems application // J. Intelligent Transportation Systems. 1993. V. 1. No. 1. P. 1–11.
5. *Гимади Э. Х., Глебов Н. И.* Математические модели и методы принятия решений. Новосибирск: Изд-во Новосиб. ун-та, 2008. 162 с.
6. *Wagner D. and Willhalm T.* Speed-up techniques for shortest-path computations // LNCS. 2007. V. 4393. P. 23–36.
7. *Goldberg A. and Harrelson C.* Computing the shortest path: A* search meets graph theory // Proc. 16th Ann. ACM-SIAM Symp. Discrete Algorithms (SODA). 2005. P. 156–165.
8. *Nejad M. M., Mashayekhy L., Chinnam R. B., and Phillips A.* Hierarchical time-dependent shortest path algorithms for vehicle routing under ITS // J. IIE Trans. 2016. V. 48. No. 2. P. 158–169.
9. *Delling D.* Time-dependent SHARC-routing // J. Algorithmica. 2011. V. 60. No. 1. P. 60–94.
10. *Delling D. and Wagner D.* Landmark-based routing in dynamic graphs // LNCS. 2007. V. 4525. P. 52–65.
11. *Goldberg A., Kaplan H., and Werneck R.* Better landmarks within reach // LNCS. 2007. V. 4525. P. 38–51.
12. `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.396.6069&rep=rep1&type=pdf` — A Landmark Algorithm for the Time-Dependent Shortest Path Problem. Tatsuya OHSHIMA. 2008.
13. *Рассел С., Норвиг П.* Искусственный интеллект. Современный подход. М.: Вильямс, 2007. 1410 с.
14. `http://www.dis.uniroma1.it/challenge9/` — DIMACS Implementation Challenge. Shortest Paths. 2006.

REFERENCES

1. *Emelichev V. A., Mel'nikov O. I., Sarvanov V. I., and Tyshkevich R. I.* Lektsii po teorii grafov [Lectures in Graph Theory]. Moscow, Librokom Publ., 2012. 390 p. (in Russian)
2. *Kormen T. Kh., Leyzerson Ch. I., Rivest R. L., and Shtayn K.* Algoritmy. Postroenie i analiz. Algoritmy. Postroyeniye i analiz. [Introduction to Algorithms]. (in Russian) Moscow, Vil'yams Publ., 2016. 1328 p.
3. *Sherali H. D., Ozbay K., and Subramanian S.* The time-dependent shortest pair of disjoint paths problem: Complexity, models, and algorithms. J. Networks, 1998, vol. 31, no. 4, pp. 259–272.
4. *Kaufman D. E. and Smith R. L.* Fastest paths in time-dependent networks for intelligent vehicle-highway systems application. J. Intelligent Transportation Systems, 1993, vol. 1, no. 1, pp. 1–11.
5. *Gimadi E. Kh. and Glebov N. I.* Matematicheskiye modeli i metody prinyatiya resheniy. [Mathematical Models and Methods of decision-making]. Novosibirsk, NSU Publ., 2008. 162 p. (in Russian)

6. *Wagner D. and Willhalm T.* Speed-up techniques for shortest-path computations. LNCS, 2007, vol. 4393, pp. 23–36.
7. *Goldberg A. and Harrelson C.* Computing the Shortest Path: A* Search Meets Graph Theory. Proc. 16th Ann. ACM-SIAM Symp. Discrete Algorithms (SODA), 2005, pp. 156–165.
8. *Nejad M. M., Mashayekhy L., Chinnam R. B., and Phillips A.* Hierarchical time-dependent shortest path algorithms for vehicle routing under ITS. J. IIE Trans., 2016, vol. 48, no. 2, pp. 158–169.
9. *Delling D.* Time-dependent SHARC-routing. J. Algorithmica, 2011, vol. 60, no. 1, pp. 60–94.
10. *Delling D. and Wagner D.* Landmark-based routing in dynamic graphs. LNCS, 2007, vol. 4525, pp. 52–65.
11. *Goldberg A., Kaplan H., and Werneck R.* Better landmarks within reach. LNCS, 2007, vol. 4525, pp. 38–51.
12. `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.396.6069&rep=rep1&type=pdf` — A Landmark Algorithm for the Time-Dependent Shortest Path Problem. Tatsuya OHSHIMA, 2008.
13. *Rassel S. and Norvig P.* Iskusstvennyy intellekt. Sovremennyy podkhod [Artificial Intelligence: A Modern Approach]. Moscow, Vil'yams Publ., 2007. 1410 p. (in Russian)
14. `http://www.dis.uniroma1.it/challenge9/` — DIMACS Implementation Challenge. Shortest Paths. 2006.

СВЕДЕНИЯ ОБ АВТОРАХ

АЛЕХИНА Марина Анатольевна — доктор физико-математических наук, профессор, проректор по научной работе Пензенского государственного технологического университета, г. Пенза. E-mail: **alekhina.marina19@yandex.ru**

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**

БАРСУКОВА Оксана Юрьевна — кандидат физико-математических наук, доцент Пензенского государственного университета, г. Пенза. E-mail: **kuzya_7@mail.ru**

БЫКОВА Валентина Владимировна — доктор физико-математических наук, доцент, профессор Сибирского федерального университета, г. Красноярск.
E-mail: **bykvalen@mail.ru**

ЕГОРУШКИН Олег Игоревич — старший преподаватель кафедры прикладной математики Сибирского государственного университета науки и технологий им. акад. М. Ф. Решетнёва, г. Красноярск. E-mail: **egorushkin.o@yandex.ru**

КОЛБАСИНА Ирина Валерьевна — аспирантка кафедры прикладной математики Сибирского государственного университета науки и технологий им. акад. М. Ф. Решетнёва, г. Красноярск. E-mail: **kabaskina@yandex.ru**

КОРЕНЕВА Алиса Михайловна — аспирантка кафедры криптологии и кибербезопасности Национального исследовательского ядерного университета «МИФИ», системный аналитик ООО «Код Безопасности», г. Москва.
E-mail: **alisa.koreneva@gmail.com**

КОСТЮК Юрий Леонидович — доктор технических наук, профессор Национального исследовательского Томского государственного университета, г. Томск.
E-mail: **kostyuk_y_l@sibmail.com**

НОВОСЕЛОВ Семен Александрович — ассистент Балтийского федерального университета им. И. Канта, г. Калининград. E-mail: **snovoselov@kantiana.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 Омского государственного университета им. Ф. М. Достоевского, г. Омск.
E-mail: **romankov48@mail.ru**

РЫБАЛОВ Александр Николаевич — кандидат физико-математических наук, доцент кафедры компьютерной математики и программирования Омского государственного университета им. Ф. М. Достоевского, г. Омск; старший научный сотрудник Института математики им. С. Л. Соболева СО РАН, г. Новосибирск.
E-mail: **alexander.rybalov@gmail.com**

САФОНОВ Константин Владимирович — доктор физико-математических наук, доцент, заведующий кафедрой прикладной математики Сибирского государственного университета науки и технологий им. акад. М. Ф. Решетнёва, г. Красноярск.
E-mail: **safonovkv@rambler.ru**

СОЛДАТЕНКО Александр Александрович — магистрант Института математики и фундаментальной информатики Сибирского федерального университета, г. Красноярск. E-mail: **glinckon@gmail.com**

ТАРКОВ Михаил Сергеевич — кандидат технических наук, доцент, старший научный сотрудник Института физики полупроводников СО РАН, г. Новосибирск.
E-mail: **tarkov@isp.nsc.ru**

Журнал «Прикладная дискретная математика» включен в перечень ВАК рецензируемых российских журналов, в которых должны быть опубликованы основные результаты диссертаций, представляемых на соискание учёной степени кандидата и доктора наук, в базы данных Web of Science (Emerging Sources Citation Index — ESCI и Russian Science Citation Index — RSCI), Scopus и MathSciNet, а также в перечень журналов, рекомендованных ФУМО ВО ИБ в качестве учебной литературы по специальности «Компьютерная безопасность».

Журнал «Прикладная дискретная математика» распространяется по подписке; его подписной индекс 38696 в объединённом каталоге «Пресса России». Полнотекстовые электронные версии вышедших номеров журнала доступны на его сайте journals.tsu.ru/pdm и на Общероссийском математическом портале www.mathnet.ru. На сайте журнала можно найти также правила подготовки рукописей статей в журнал.

Тематика публикаций журнала:

- *Теоретические основы прикладной дискретной математики*
- *Математические методы криптографии*
- *Математические методы стеганографии*
- *Математические основы компьютерной безопасности*
- *Математические основы надёжности вычислительных и управляющих систем*
- *Прикладная теория кодирования*
- *Прикладная теория автоматов*
- *Прикладная теория графов*
- *Логическое проектирование дискретных автоматов*
- *Математические основы информатики и программирования*
- *Вычислительные методы в дискретной математике*
- *Дискретные модели реальных процессов*
- *Математические основы интеллектуальных систем*
- *Исторические очерки по дискретной математике и её приложениям*